

GTLL_MOUSE
 ID GTLL_MOUSE STANDARD; PRT; 558 AA.
 AC Q9JG61;
 DT 01-OCT-2004 (Rel. 45, Created)
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Putative polypeptide N-acetylglucosaminyltransferase-like protein 1
 DE (EC 2.4.1.41) (Protein-UDP acetylglucosaminyltransferase-like
 DE protein 1) (UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase-
 DE like protein 1) (Polypeptide GalNAc transferase-like protein 1)
 DE (GalNAc-T-like protein 1) (pp-GaNTase-like protein 1).
 GN Name=Galnt1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Osada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
 RA Hashimoto K.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP TISSUE SPECIFICITY.
 RP PubMed=15018805;
 RA Nelson P.A., Sutcliffe J.G., Thomas E.A.;
 RA "A new UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase mRNA
 RT exhibits predominant expression in the hypothalamus, thalamus and
 RT amygdala of mouse forebrain."
 RL Gene Expr. Patterns 1:95-99(2002).
 CC -!- FUNCTION: May catalyze the initial reaction in O-linked
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-
 CC galactosamine residue to a serine or threonine residue on the
 CC protein receptor (By similarity).
 CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
 CC -!- COFACTOR: Manganese and calcium (By similarity).
 CC -!- PATHWAY: Glycosylation.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 CC similarity).
 CC -!- TISSUE SPECIFICITY: In the CNS, it is predominantly expressed in
 CC several distinct hypothalamic, thalamic and amygdaloid nuclei. The
 CC most abundant level of expression is in the paraventricular,
 CC ventromedial and arcuate nuclei of the hypothalamus, the
 CC anterodorsal and parafascicular nuclei of the thalamus and the
 CC central, basomedial and medial nuclei of the amygdala. Also
 CC expressed in cerebral cortex, lateral septum, habenula and
 CC hippocampus.
 CC -!- DOMAIN: There are two conserved domains in the glycosyltransferase
 CC region: the N-terminal domain (domain A, also called GT1 motif),
 CC which is probably involved in manganese coordination and substrate
 CC binding and the C-terminal domain (domain B, also called
 CC Gal/GalNAc-T motif), which is probably involved in catalytic
 CC reaction and UDP-Gal binding (By similarity).
 CC -!- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
 CC contributes to the glycopeptide specificity (By similarity).
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
 CC subfamily.
 CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -!- CAUTION: Was originally (Ref.2) termed Galnt10/pp-GaNTase 10.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 EMBL; AB045325; BAA97985.1; -
 MGD; MG1:1917754; Galnt11.
 DR InterPro; IPR001173; Glyco trans 2.
 DR InterPro; IPR008997; RicinB_like.

DR InterPro; IPR000772; Ricin_B_lectin.
 DR Pfam; PF00535; Glycos transf_2; 1.
 DR Pfam; PF03652; Ricin_B_lectin; 2.
 DR SMART; SM00458; RICIN; 1.
 DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
 KW Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;
 KW Signal-anchor; Transferase; Transmembrane.
 FT DOMAIN 1 6
 FT TRANSMEM 7 26
 FT Cytoplasmic (Potential).
 FT Signal-anchor for type II membrane
 FT protein (Potential).
 FT Lumenal (Potential).
 FT DOMAIN 27 558
 FT DOMAIN 122 227
 FT DOMAIN 286 348
 FT DOMAIN 428 555
 FT DISULFID 441 460
 FT DISULFID 486 506
 FT DISULFID 530 543
 FT By similarity.
 FT By similarity.
 SQ SEQUENCE 558 AA; 63019 MW; 2FE9B0CAD13FC8AE CRC64;
 Query Match 55.4%; Score 128.5; DB 1; Length 558;
 Best Local Similarity 55.4%; Pred. No. 6.9e-09;
 Matches 25; Conservative 6; Mismatches 11; Indels 3; Gaps 2;
 QY 1 CGGRMEDIPCSRCHYKVPYKVPAGVSLA--RNLRVAD-WM 42
 DB 331 CGGSLEIVPCSRVGHVFRKHYPNPEGNALTYIRNKTAEVWM 375
 RESULT 10
 GLT2_MOUSE
 ID GLT2_MOUSE STANDARD; PRT; 570 AA.
 AC Q6PB93; Q7TS15; Q9SL27; Q922K5; Q99ME1;
 DT 01-OCT-2004 (Rel. 45, Created)
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Polypeptide N-acetylglucosaminyltransferase 2 (EC 2.4.1.41)
 DE (Protein-UDP acetylglucosaminyltransferase 2) (UDP-
 DE GalNAc:polypeptide N-acetylglucosaminyltransferase 2) (Polypeptide
 DE GalNAc transferase 2) (GalNAc-T2) (pp-GaNTase 2) [Contains:
 DE Polypeptide N-acetylglucosaminyltransferase 2, soluble form].
 GN Name=Galnt2;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Colon adenocarcinoma;
 RA Miyahara N., Kanoh A., Irimura T.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RP STRAIN=CS7BL/6J; TISSUE=Adipose tissue;
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schorbach C., Gojobori T.,
 RA Balzarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schirl L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragoti T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gattiboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustinchik S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Karai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konegaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons B.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perle G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sanderlin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,

```
CC subfamily.
CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute Of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/embl/
CC or send an email to licenseseis@sib.ch).
CC -----
CC EMBL; AF348968; AAX37548.1; -
CC EMBL; AK046567; BAC32790.1; -
CC EMBL; BC007172; AAHC07172.1; -
CC EMBL; BC053063; AAH53063.1; -
CC EMBL; BC059818; AAH59818.1; -
CC HSSP; P26514; 1KNM.
CC MGD; MG1:894694; Galnt2.
CC InterPro; IPR001173; Glyco_trans_2.
CC InterPro; IPR008997; RicinB_like.
CC InterPro; IPR000772; Ricin_B_lectin.
CC Pfam; PF00535; Glycos_transf_2; 1.
CC Pfam; PF00652; Ricin_B_lectin; 3.
CC SMART; SM00458; RICIN-B_1.
CC PROSITE; PS02331; RICIN_B_LECTIN; 1.
CC KW Alternative splicing; Calcium; Glycosyltransferase; Golgi stack;
CC Lactin; Manganese; Signal-anchor; transferase; Transmembrane.
CC FT PROPEP 1 50 Removed in soluble polypeptide N-
FT acetyl-galactosaminyltransferase 2 (By
FT similarity).
FT CHAIN 51 570 Polypeptide N-
FT acetyl-galactosaminyltransferase 2,
FT soluble form.
FT DOMAIN 1 6 Cytoplasmic (Potential).
FT TRANSMEM 7 24 Signal-anchor for type II membrane
FT protein (Potential).
FT Lumenal (Potential).
FT DOMAIN 25 570 Catalytic subdomain A.
FT DOMAIN 134 239 Catalytic subdomain B.
FT DOMAIN 299 361 Ricin B-type lectin.
FT DOMAIN 442 565 By similarity.
FT SITE 515 515 Not glycosylated (by similarity).
FT DISULFID 455 472 By similarity.
FT DISULFID 495 512 By similarity.
FT DISULFID 538 554 By similarity.
FT VARSPLIC 1 41 MRERSMLCLFALLWLVLGIAYVYSGGSALLAAGGGGAGRK
FT -> MALENPO (in isoform 2).
FT /FTid=VSP_011201.
FT R -> L (in Ref. 1).
FT C -> Y (in Ref. 2).
FT DSR -> NSK (in Ref. 1).
FT F -> V (in Ref. 1).
FT SEQUENCE 570 AA; 64514 MW; 90D5DC02C85A8EA CRC64;

Query Match 55.4%; Score 128.5; DB 1; Length 570;
Best Local Similarity 55.6%; Pred.No. 7.1e-09;
Matches 25; Conservative 6; Mismatches 11; Indels 3; Gaps 2;

QY 1 CGGRVEDIPCSRGVGHYRKVPVKVP--AGVSLARLNKEVAD-WM 42
DB 344 CGGSLEIIPCVRGVGHVFVKQHPIFFCGSGTVFAIRTRAAEAVWM 388
|||||:|||||||::|||:|||||:|||||
CGRVEDIPCSRGVGHYRKVPVKVP--AGVSLARLNKEVAD-WM 42
CGGSLEIIPCVRGVGHVFVKQHPIFFCGSGTVFAIRTRAAEAVWM 388

RESULT 11
AAHS9818
ID AAHS9818 PRELIMINARY; PRT; 570 AA.
AC AAHS9818;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DE 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE UDP-N-acetyl-alpha-D-galactosamine; polypeptide
DE N-acetyl-galactosaminyltransferase 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M.J., Udén T.B., Toshikiyuki S., Carninci P., Scheetz T.E.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC059818; AAH59818.1; -;
 KW transferase.
 SQ SEQUENCE 570 AA; 64514 MW; 90D5DC02C85A8EEA CRC64;
 Query Match 55.4%; Score 128.5; DB 2; Length 570;
 Best Local Similarity 55.6%; Pred. No. 7.1e-09;
 Matches 25; Conservative 6; Mismatches 11; Indels 3; Gaps 2;
 QY 1 CGGRMEDIPCSRGVHRYKVPYKVP--AGVSLARNLKRVD-WM 42
 DB 344 CGSLEIPCSRGVHRYKVPYKVP--AGVSLARNLKRVD-WM 388
 RESULT 12
 ID GLT2_HUMAN STANDARD; PRT; 571 AA.
 AC Q10471; OSNPY4;
 DT 01-OCT-2004 (Rel. 45, Created)
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Polypeptide N-acetylgalactosaminyltransferase 2 (EC 2.4.1.41)
 DE (Protein-UDP N-acetylgalactosaminyltransferase 2)
 DE GalNac:polypeptide N-acetylgalactosaminyltransferase 2 (Polypeptide
 DE GalNac transferase 2) (GalNac-T2) (pp-GalNTase 2) [Contains:
 DE Polypeptide N-acetylgalactosaminyltransferase 2, soluble form].
 GN Name=GLNT2;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.; SEQUENCE OF 52-93; 104-120; 170-182; 193-218;
 RP 303-323; 421-424 AND 510-546, ENZYME ACTIVITY, AND TISSUE SPECIFICITY.
 RC TISSUE=Gastric carcinomas;
 RX MEDLINE=96025800; PubMed=7592619;
 RA White T., Bennett E., Paul E., Takio K., Sorensen T., Bonding N.,
 RA Clausen H.;
 RT "Purification and cDNA cloning of a human UDP-GalNac:polypeptide N-
 RT acetyl-galactosaminyltransferase.";
 RL J. Biol. Chem. 270:24156-24165 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Donnelly S.;
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M.J., Udén T.B., Toshikiyuki S., Carninci P., Scheetz T.E.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]
 RP FUNCTION.
 EX PubMed=9295285;
 RA Wandall H.H., Hassan H., Mirgorodskaya E., Kristensen A.K.,
 RA Roepstorff P., Bennett E.P., Nielsen P.A., Hollingsworth M.A.,
 RA Burchell J., Taylor-Papadimitriou J., Clausen H.;
 RT "Substrate specificities of three members of the human UDP-N-acetyl-
 RT alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase
 RT family, GalNac-T1, -T2 and -T3";
 RL J. Biol. Chem. 272:23503-23514 (1997).
 RN [5]
 RP SUBCELLULAR LOCATION.
 EX PubMed=9394011;
 RA Roettger S., White J., Wandall H.H., Olivo J.-C., Stark A.,
 RA Bennett E.P., Whitehouse C., Berger E.G., Clausen H., Nilsson T.;
 RT "Localization of three human polypeptide GalNac-transferases in HeLa
 RT cells suggests initiation of O-linked glycosylation throughout the
 RT Golgi apparatus.";
 RL J. Cell Sci. 111:45-60 (1998).
 RN [6]
 RP FUNCTION.
 EX PubMed=12438318; DOI=10.1074/jbc.M211097200;
 RA Iwasaki H., Zhang Y., Tachibana K., Gorch M., Kikuchi N., Kwon Y.-D.,
 RA Toyayachi A., Kudo T., Kubota T., Marimatsu H.;
 RT "Initiation of O-glycan synthesis in IgA1 hinge region is determined
 RT by a single enzyme, UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-
 RT acetyl-galactosaminyltransferase 2.";
 RL J. Biol. Chem. 278:5613-5621 (2003).
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-
 CC galactosamine residue to a serine or threonine residue on the
 CC protein receptor. Has a broad spectrum of substrates for peptides
 CC such as EA2, Muc5AC, Muc1b, Muc1c. Probably involved in O-linked
 CC glycosylation of the immunoglobulin A1 (IgA1) hinge region.
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
 CC -1- COFACTOR: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides
 CC preferentially in the trans and medial parts of the Golgi stack. A
 CC secreted form also exists.
 CC -1- TISSUE SPECIFICITY: Widely expressed.
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
 CC region: the N-terminal domain (domain A, also called 571 motif),
 CC which is probably involved in manganese coordination and substrate
 CC binding and the C-terminal domain (domain B, also called
 CC Gal/GalNac-T motif), which is probably involved in catalytic

reaction and UDP-Gal binding (By similarity).
CC -i- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
CC contributes to the glycopeptide specificity (By similarity).
CC -i- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
CC subfamily.
CC -i- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; X85019; CAA59381.1; -;
DR EMBL; AL078646; CAC00585.1; -;
DR EMBL; AL136988; -; NOT ANNOTATED_CDS.
DR EMBL; AL117349; -; NOT ANNOTATED_CDS.
DR EMBL; BC041120; AA41120.1; -;
DR PIR; I37405; I37405.
DR HSP; P26514; IKNM.
DR Genew; HGNC:4124; GALNT2.
DR MIM; 602274; -;
DR GO; GO:0004653; P:polypeptide N-acetylgalactosaminyltransferase. . . TAS.
DR GO; GO:0006493; P:O-linked glycosylation; TAS.
DR InterPro; IPR001173; Glyco trans 2.
DR InterPro; IPR008997; RicinB like.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00535; Glycos transf 2; 1.
DR Pfam; PF00652; Ricin B lectin; 3.
DR SMART; SM00458; RICIN_1
DR PROSITE; PS00231; RICIN_B_LECTIN; 1.
DR Calcium; Direct protein sequencing; Glycosyltransferase; Golgi stack;
KW Lectin; Manganese; Polymorphism; Signal-anchor; Transferase;
KW Transmembrane.
FT PROPEP 1 51 Removed in soluble polypeptide N-
FT acetyl-galactosaminyltransferase 2.
FT CHAIN 52 571 acetyl-galactosaminyltransferase 2,
FT soluble form.
FT DOMAIN 1 6 Cytoplasmic (Potential).
FT TRANSMEM 7 24 Signal-anchor for type II membrane
FT protein (Potential).
FT DOMAIN 25 571 Luminal (Potential).
FT DOMAIN 135 240 Catalytic subdomain A.
FT DOMAIN 300 362 Catalytic subdomain B.
FT DOMAIN 443 566 Ricin B-type lectin.
FT SITE 516 516 Not glycosylated (Probable).
FT SITE 456 473 By similarity.
FT DISULFID 496 513 By similarity.
FT DISULFID 539 555 By similarity.
FT VARIANT 554 554 V -> M (in dbSNP:2273970).
FT /Fid=VAR_013575.
FT CONFLICT 70 70 T -> G (in Ref. 1; AA sequence).
FT CONFLICT 78 78 W -> D (in Ref. 1; AA sequence).
FT CONFLICT 93 93 R -> D (in Ref. 1; AA sequence).
FT CONFLICT 210 210 R -> W (in Ref. 1; AA sequence).
FT CONFLICT 290 291 RR -> SC (in Ref. 1; AA sequence).
FT CONFLICT 293 293 R -> Q (in Ref. 1; AA sequence).
FT CONFLICT 300 300 P -> H (in Ref. 1; AA sequence).
FT CONFLICT 300 300 W -> A (in Ref. 1; AA sequence).
FT CONFLICT 522 522 H -> M (in Ref. 1; AA sequence).
FT CONFLICT 533 533 H -> M (in Ref. 1; AA sequence).
SQ SEQUENCE 571 AA; 64732 MW; D9A0F5D17C55BAF2 CRC64;
Query Match 55.4%; Score 128.5; DB 1; Length 571;
Best Local Similarity 55.6%; Pred. No. 7.1e-09;
Matches 25; Conservative 6; Mismatches 11; Indels 3; Gaps 2;
Qy 1 CGGMEIPCSRUGHYKVPYKVP--AGVSLRNKRKRVAD-WM 42
DB 345 CGGSLIIPCSRUGHYKVPYKVPYKVP--PSAKTYDFVAKHKRVAVWVM 794

RESULT 13
Q7QDR0
ID Q7QDR0 PRELIMINARY; PRT; 1003 AA.
AC Q7QDR0;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AGCP10665 (Fragment).
GN Name=agCG47419; ORFNames=ENSANGG00000013497;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Anopheles.
OC NCSI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RL Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -i- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008849; EAA07231.1; -;
DR InterPro; IPR001173; Glyco trans 2.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00535; Glycos transf 2; 2.
DR Pfam; PF00652; Ricin B lectin; 4.
DR PROSITE; PS50231; RICIN_B_LECTIN; 2.
FT NON_TER 1 1
SQ SEQUENCE 1003 AA; 115923 MW; 753EAS0F567A4E13 CRC64;
Query Match 55.2%; Score 128; DB 2; Length 1003;
Best Local Similarity 58.7%; Pred. No. 1.5e-08;
Matches 27; Conservative 6; Mismatches 7; Indels 6; Gaps 3;
Qy 1 CGGMEIPCSRUGHYKVPYKVPAGVS--LARNLKRKRVAD-WM 42
DB 751 CGGMEIPCSRUGHYKVPYKVPYKVP--PSAKTYDFVAKHKRVAVWVM 794

RESULT 14
AAS64620
ID AAS64620 PRELIMINARY; PRT; 590 AA.
AC AAS64620;
DT 01-APR-2004 (TrEMBLrel. 27, Created)
DT 01-APR-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-APR-2004 (TrEMBLrel. 27, Last annotation update)
DE CG3254-PB.
GN CG3254.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 12.7807 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-33
Perfect score: 203
Sequence: 1 ALSWTMEMEMLLARFRTGDLHDSVHLCAHP 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pap:*
2: /cgn2_6/prodata/1/iaa/5B COMB.pap:*
3: /cgn2_6/prodata/1/iaa/6A COMB.pap:*
4: /cgn2_6/prodata/1/iaa/6B COMB.pap:*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pap:*
6: /cgn2_6/prodata/1/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	26.6	106	4	US-09-084-303B-2
2	53	26.1	416	4	Sequence 33025, A
3	51	25.1	669	4	Sequence 25877, A
4	49	24.1	283	4	Sequence 8529, Ap
5	49	24.1	1297	4	Sequence 6373, Ap
6	48.5	23.9	575	4	Sequence 32986, A
7	48	23.6	172	4	Sequence 16, Appl
8	48	23.6	288	4	Sequence 2068, Ap
9	48	23.6	353	4	Sequence 5230, Ap
10	47.5	23.4	653	4	Sequence 20015, A
11	47	23.2	66	4	Sequence 36931, A
12	47	23.2	66	4	Sequence 52148, A
13	47	23.2	273	4	Sequence 12, Appl
14	47	23.2	273	4	Sequence 14, Appl
15	47	23.2	273	4	Sequence 16, Appl
16	47	23.2	273	4	Sequence 20, Appl
17	47	23.2	273	4	Sequence 22, Appl
18	47	23.2	318	4	Sequence 18, Appl
19	47	23.2	435	4	Sequence 31215, A
20	46.5	22.9	1201	3	Sequence 2, Appl1
21	46	22.7	206	4	Sequence 26202, A
22	46	22.7	317	4	Sequence 42690, A
23	46	22.7	334	4	Sequence 32245, A
24	46	22.7	385	4	Sequence 7438, Ap
25	46	22.7	412	4	Sequence 23193, A
26	46	22.7	554	4	Sequence 23106, A
27	46	22.7	569	4	Sequence 106, App

28 46 22.7 680 4 US-09-489-039A-10982 Sequence 10982, A
29 45.5 22.4 358 4 US-09-759-875-8 Sequence 9, Appli
30 45.5 22.4 360 3 US-09-509-902A-9 Sequence 9, Appli
31 45.5 22.4 412 1 US-08-102-863-11 Sequence 11, Appl
32 45.5 22.4 412 5 PCT-US92-10885-11 Sequence 11, Appl
33 45.5 22.4 714 4 US-09-252-991A-29806 Sequence 29806, A
34 45 22.2 79 4 US-09-134-000C-5615 Sequence 5615, Ap
35 45 22.2 174 4 US-09-270-767-35433 Sequence 35433, A
36 45 22.2 174 4 US-09-270-767-50650 Sequence 50650, A
37 45 22.2 262 1 US-08-738-944-50 Sequence 50, Appl
38 45 22.2 262 3 US-09-263-352-40 Sequence 40, Appl
39 45 22.2 265 4 US-09-252-991A-28073 Sequence 28073, A
40 45 22.2 398 4 US-09-252-991A-17379 Sequence 17379, A
41 45 22.2 465 3 US-09-292-097-15 Sequence 15, Appl
42 45 22.2 465 4 US-09-933-561-15 Sequence 15, Appl
43 45 22.2 494 4 US-03-252-991A-17295 Sequence 17295, A
44 45 22.2 496 4 US-09-252-991A-22592 Sequence 22592, A
45 45 22.2 596 2 US-08-836-620A-16 Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-084-303B-2
; Sequence 2, Application US/09084303B
; Patent No. 6627746
; GENERAL INFORMATION:
; APPLICANT: Doberstein, Stephen
; APPLICANT: Reddy, Bindu
; APPLICANT: Platt, Darren
; TITLE OF INVENTION: FERGUSON, Kimberly
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS OF C. ELEGANS INSULIN-LIKE GENES AND US
; FILE REFERENCE: 7326-069-999
; CURRENT APPLICATION NUMBER: US/09/084,303B
; CURRENT FILING DATE: 1998-05-26
; NUMBER OF SEQ ID NOS: 302
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-084-303B-2

Query Match 26.6%; Score 54; DB 4; Length 106;
Best Local Similarity 61.1%; Pred. No. 1.3;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 5 IEMTEMLLARRRTP 22
Db 36 IEMTEMLLARRRTP 53

RESULT 2
US-09-252-991A-33025
; Sequence 33025, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 33025
; LENGTH: 416
; TYPE: PRT

RESULT 10
US-09-248-796A-20015

Query Match 23.2%; Score 47; DB 4; Length 66;
Best Local Similarity 37.0%; Pred. No. 8.5;
Matches 10; Conservative 1; Mismatches 8; Indels 8; Gaps 1;
Qy 19 RRTPGDLHLD-----HSVHLCAHP 37
Db 32 RHGPNVYLDGGLRLCSHSHVQCTHP 58

RESULT 13
US-09-396-149-12
; Sequence 12, Application US/09396149
; Patent No. 6538176
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Replication Protein A and Use
; FILE REFERENCE: 5718-59
; CURRENT APPLICATION NUMBER: US/09/396,149
; CURRENT FILING DATE: 1999-09-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Zea mays
US-09-396-149-12

Query Match 23.2%; Score 47; DB 4; Length 273;
Best Local Similarity 30.8%; Pred. No. 45;
Matches 8; Conservative 6; Mismatches 12; Indels 0; Gaps 0;
Qy 8 DTEMEMLLARFRRTPGDLHLDHSVHL 33
Db 208 DTDLHTQVLNFFNEPANLESEHGHHV 233

RESULT 14
US-09-396-149-14
; Sequence 14, Application US/09396149
; Patent No. 6538176
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Replication Protein A and Use
; FILE REFERENCE: 5718-59
; CURRENT APPLICATION NUMBER: US/09/396,149
; CURRENT FILING DATE: 1999-09-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Zea mays
US-09-396-149-14

Query Match 23.2%; Score 47; DB 4; Length 273;
Best Local Similarity 30.8%; Pred. No. 45;
Matches 8; Conservative 6; Mismatches 12; Indels 0; Gaps 0;
Qy 8 DTEMEMLLARFRRTPGDLHLDHSVHL 33
Db 208 DTDLHTQVLNFFNEPANLESEHGHHV 233

RESULT 15
US-09-396-149-16
; Sequence 16, Application US/09396149
; Patent No. 6538176
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Replication Protein A and Use
; FILE REFERENCE: 5718-59
; CURRENT APPLICATION NUMBER: US/09/396,149

; CURRENT FILING DATE: 1999-09-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Zea mays
US-09-396-149-16

Query Match 23.2%; Score 47; DB 4; Length 273;
Best Local Similarity 30.8%; Pred. No. 45;
Matches 8; Conservative 6; Mismatches 12; Indels 0; Gaps 0;
Qy 8 DTEMEMLLARFRRTPGDLHLDHSVHL 33
Db 208 DTDLHTQVLNFFNEPANLESEHGHHV 233

Search completed: November 10, 2004, 13:44:05
Job time : 12.7807 secs

RESULT 2
US-10-425-115-357905
; Sequence 357905, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:

```

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 357905
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(141)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_8957C.1.pap
US-10-425-115-357905

```

```

Query Match      30.5%; Score 62; DB 17; Length 141;
Best Local Similarity 37.8%; Pred. No. 0.7;
Matches 14; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 ALSWIENDTEMEMLLARFRTPGDLHLDHSHVLCAP 37
      |||::|||::|||::|||::|||::|||::|||
DB 96 AFSAVLLDSYELLRLXVLTQGTGTFPLDRRTCCRP 132

```

```

RESULT 3
US-10-437-963-167475
; Sequence 167475, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 167475
; LENGTH: 1703
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_66083C.1.pap
US-10-437-963-167475

```

```

Query Match      27.8%; Score 56.5; DB 16; Length 1703;
Best Local Similarity 46.9%; Pred. No. 74;
Matches 15; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

```

```

QY 10 EMEMLLARFRTPG-----DLHLDHSHVLCAP 36
      |||::|||::|||::|||::|||::|||
DB 567 EQRLLARIPRTGRLYMLDINLAREVCLAHH 598

```

```

RESULT 4
US-10-437-963-122236
; Sequence 122236, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

```

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122236
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_25183C.1.pap
US-10-437-963-122236

```

```

Query Match      27.3%; Score 55.5; DB 16; Length 213;
Best Local Similarity 43.8%; Pred. No. 10;
Matches 14; Conservative 4; Mismatches 9; Indels 5; Gaps 2;

```

```

QY 5 IEMDTEMEMLLARFRTPGDLHLDHSHVLCAP 36
      |||::|||::|||::|||::|||::|||
DB 119 IEMD-QVAMAIIDR----PIDRHLHLQRTCAH 145

```

```

RESULT 5
US-10-437-963-167473
; Sequence 167473, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 167473
; LENGTH: 1136
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_66081C.1.pap
US-10-437-963-167473

```

```

Query Match      27.3%; Score 55.5; DB 16; Length 1136;
Best Local Similarity 46.9%; Pred. No. 66;
Matches 15; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

```

```

QY 10 EMEMLLARFRTPG-----DLHLDHSHVLCAP 36
      |||::|||::|||::|||::|||::|||
DB 473 EQRLLARIPRTGRLYMLDINLAREVCLAHH 504

```

```

RESULT 6
US-09-205-658-118
; Sequence 118, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott

```

RESULT 8
US-10-339-740-2
; Sequence 2, Application US/10339740
; Publication No. US20030187246A1
; GENERAL INFORMATION:
; APPLICANT: Doberstein, Stephen
; APPLICANT: Reddy, Bindu
; APPLICANT: Platt, Darren

```

RESULT 10
US-10-437-963-104186
; Sequence 104186, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbatov, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid M
; TITLE OF INVENTION: Plants and Uses Th
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,9
; CURRENT FILING DATE: 2003-05-14

```

```
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 104186
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_101544C.1.pep
US-10-437-963-104186

Query Match      26.1%; Score 53.5; DB 16; Length 1431;
Best Local Similarity 45.2%; Pred. No. 1.7e+02;
Matches 14; Conservative 3; Mismatches 9; Indels 5; Gaps 1;

QY 10 EMEMLLARFRRTPG-----DLHLDHSHVLC 35
Db 474 EQHELLARITTPGRVLVDINLARPYLMA 504

RESULT 11
US-10-424-599-157136
; Sequence 157136, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 157136
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112915C.1.pep
US-10-424-599-157136

Query Match      26.1%; Score 53; DB 15; Length 50;
Best Local Similarity 41.2%; Pred. No. 4.5;
Matches 14; Conservative 3; Mismatches 9; Indels 8; Gaps 1;

QY 2 LSWIEMTEMLLARFRRTPGDLHLDHSHVLC 35
Db 20 LFWIIMDTHPECLLSI-----LDYWLGLCA 45

RESULT 12
US-10-425-115-350053
; Sequence 350053, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 39-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 350053
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_82421C.1.pep
US-10-425-115-350053

; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 104186
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_101544C.1.pep
US-10-437-963-104186

Query Match      26.1%; Score 53; DB 16; Length 1431;
Best Local Similarity 45.2%; Pred. No. 1.7e+02;
Matches 14; Conservative 3; Mismatches 9; Indels 5; Gaps 1;

QY 10 EMEMLLARFRRTPG-----DLHLDHSHVLC 35
Db 474 EQHELLARITTPGRVLVDINLARPYLMA 504

RESULT 11
US-10-424-599-157136
; Sequence 157136, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 157136
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112915C.1.pep
US-10-424-599-157136

Query Match      26.1%; Score 53; DB 15; Length 50;
Best Local Similarity 41.2%; Pred. No. 4.5;
Matches 14; Conservative 3; Mismatches 9; Indels 8; Gaps 1;

QY 2 LSWIEMTEMLLARFRRTPGDLHLDHSHVLC 35
Db 20 LFWIIMDTHPECLLSI-----LDYWLGLCA 45

RESULT 12
US-10-425-115-350053
; Sequence 350053, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 39-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 350053
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_82421C.1.pep
US-10-425-115-350053

Query Match      26.1%; Score 53; DB 17; Length 59;
Best Local Similarity 52.4%; Pred. No. 5.5;
Matches 11; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 21 TPGDLHLDHSHVH---LCAHP 37
Db 19 TPHLLHLNNSYHGVGAVICAH 39

RESULT 13
US-10-369-493-3101
; Sequence 3101, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3101
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(338)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3101

Query Match      26.1%; Score 53; DB 14; Length 338;
Best Local Similarity 25.0%; Pred. No. 39;
Matches 13; Conservative 6; Mismatches 11; Indels 22; Gaps 1;

QY 5 IEMDTEMLLARFRRTPGD-----LHLDHSHVLC 34
Db 242 IELDAQYDDILGRFORPWHSHFINAENQRFVSNBAIDFLDKLRYDHNVRFC 293

RESULT 14
US-09-764-847-727
; Sequence 727, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 727
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-09-764-847-727

Query Match      25.6%; Score 52; DB 9; Length 51;
Best Local Similarity 27.3%; Pred. No. 6.5;
Matches 9; Conservative 5; Mismatches 19; Indels 0; Gaps 0;

QY 4 WIEMDTEMLLARFRRTPGDLHLDHSHVLC 36
Db 4 WHECKMKLVTLFKKLKSLPATHTHTHTTCSH 36
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 8.24352 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-33

Perfect score: 203

Sequence: 1 ALSWIEMTEMLLARFRTPGDLHLDHSHVLCARF 37

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:*

1: Piri:*

2: Piri2:*

3: Piri3:*

4: Piri4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	29.6	348	2 G83976	cytochrome caa3 ox
2	55.5	27.3	1270	2 T22615	hypothetical prote
3	55	27.1	199	2 S16063	acp-22 protein - y
4	55	27.1	199	2 S32224	acp-22 protein - y
5	54	26.6	106	2 T27989	hypothetical prote
6	52.5	25.9	886	2 T35469	probable Atp /GTP-
7	52	25.6	159	2 AF1372	proteins involved
8	52	25.6	159	2 AD1742	protein involved i
9	52	25.6	342	2 A45335	cytochrome-c oxida
10	52	25.6	599	2 E36792	hypothetical prote
11	51.5	25.4	141	2 T51959	hypothetical prote
12	51.5	25.4	211	2 S26845	dtmP kinase (SC 2.
13	51	25.1	241	2 E69392	glucose-1-phosphat
14	51	25.1	363	1 C55973	transcription fact
15	51	25.1	448	2 A71347	hypothetical prote
16	51	25.1	494	2 D83472	probable Glycosyl
17	51	25.1	770	2 S60676	cellobiose oxidase
18	50	24.6	343	2 T29819	hypothetical prote
19	50	24.6	464	2 A28569	alpha-methylidopa-h
20	50	24.6	513	2 F85095	hypothetical prote
21	50	24.6	604	2 T49577	hypothetical prote
22	49.5	24.4	242	2 C82508	hypothetical prote
23	49.5	24.4	641	1 JVBPAL	DNA-packaging prot
24	49.5	24.4	641	2 H85689	terminase large su
25	49.5	24.4	641	2 F90832	terminase large su
26	49.5	24.4	641	2 F90831	probable terminase
27	49.5	24.4	1276	2 T18526	SREBP cleavage act
28	49	24.1	158	2 C83814	Na+/H+ antiporter
29	49	24.1	167	2 T36325	hypothetical prote

RESULT 1
G83976
cytochrome caa3 oxidase (subunit II) ctaC [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: G83976
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, N.; Masui, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G83976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-348 <STO>
A:Cross-references: UNIPROT:Q9K9N0; GB:AP001516; GB:BA000004; NID:G10175192; PIDN:BAB063-
A:Experimental source: strain C-125
C:Genetics:
A:Gene: ctaC
C:Superfamily: cytochrome-c oxidase chain II, cytochrome c containing; cytochrome c6 hom
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; membrane-associated con
F:209,211,213,217/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:211/Binding site: magnesium (Glu) (shared with chain I) #status predicted
F:267,270/Binding site: heme (Cys) (covalent) #status predicted
F:271/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match 29.6%; Score 60; DB 2; Length 348;
Best Local Similarity 60.0%; Pred. No. 1;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 13 MLARFRTPGDLHLDHSHV 32
DB 66 ILARFRKPGDDELPKQV 85

RESULT 2
T22615
hypothetical protein T01B8.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22615; T24303
R:McMurray, A.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z19589
A:Accession: T22615
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1270 <WIL>
A:Cross-references: UNIPROT:Q20749; EMBL:Z48583; PIDN:CAA88473.1; GSPDB:GN000020; CESP:TO1
A:Experimental source: clone F54B3
R:McMurray, A.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z19871

30 49 24.1 251 2 A97640 hypothetical prote
31 49 24.1 251 2 A92863 conserved hypothet
32 49 24.1 339 2 T34406 hypothetical prote
33 49 24.1 423 2 T09338 DnaU-like protein
34 49 24.1 545 2 S67621 hypothetical prote
35 49 24.1 643 2 H83873 ATP-dependent heli
36 49 24.1 776 2 S67053 probable membrane
37 48.5 23.9 277 2 G69158 conserved hypothet
38 48.5 23.9 400 2 T44805 malonyl-CoA transa
39 48.5 23.9 425 2 E83023 3-deoxy-D-manno-oc
40 48.5 23.9 459 2 T15115 NADH2 dehydrogenas
41 48 23.6 352 2 T12515 hypothetical prote
42 48 23.6 478 2 A25407 interferon-induced
43 48 23.6 560 2 E70814 probable pdc prote
44 48 23.6 1106 2 C87128 probable transcrip
45 48 23.6 1191 2 S85068 DNA-directed RNA p

ALIGNMENTS

RESULT 5
T37000

A;Cross-references: UNIPROT:Q8Y4Q2; GB:NC_003210; PIDN:CAD00460.1; PID:g16411870; GSPDB:
A;Experimental source: strain EGD-e
C;Genetics:
C;Superfamily: Pyrococcus abyssi hypothetical protein PAB1401

Query Match 25.6%; Score 52; DB 2; Length 159;
Best Local Similarity 33.3%; Pred. No. 6.1; Indels 10; Gaps 1;
Matches 13; Conservative 5; Mismatches 11;

Qy 5 IEMDTEMEMLLARFRTPGDLHLD-----HSVHL 33
Db 97 LETDWEVTMLALLITLTPGTLSDISDDYKAIYVHSLV 135

RESULT 8
AD1742
protein involved in resistance to cholate and to NA(+) and in pH homeostasis homolog lin
C;Species: *Listeria innocua*
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AD1742
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecher,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ck, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland,
A.; Title: Comparative genomics of *Listeria* species
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AD1742
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-159 <GLA>
A;Cross-references: UNIPROT:Q928Q1; GB:AL592022; PIDN:CAC97708.1; PID:g164115003; GSPDB:C
A;Experimental source: strain Clp11262
C;Genetics:
C;Gene: lin2481
C;Superfamily: *Pyrococcus abyssi* hypothetical protein PAB1401

Query Match 25.6%; Score 52; DB 2; Length 159;
Best Local Similarity 33.3%; Pred. No. 6.1; Indels 10; Gaps 1;
Matches 13; Conservative 5; Mismatches 11;

Qy 5 IEMDTEMEMLLARFRTPGDLHLD-----HSVHL 33
Db 97 LETDWEVTMLALLITLTPGTLSDISDDYKAIYVHSLV 135

RESULT 9
C45335
Cytochrome-c oxidase (EC 1.9.3.1) chain II - *Bacillus firmus*
C;Species: *Bacillus firmus*
C;Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 16-Aug-2004
C;Accession: C45335; I39821
R;Quirk, P.G.; Hicks, D.B.; Krulwich, T.A.
J. Biol. Chem. 268, 678-685, 1993
A;Title: Cloning of the cta operon from alkaliphilic *Bacillus firmus* OF4 and characteriz
A;Reference number: A45335; MUID:93107080; PMID:7678007
A;Accession: C45335
A;Molecule type: DNA; protein
A;Residues: 1-342 <QUI>
A;Cross-references: GB:M94110; NID:G142782; PIDN:AAA22364.1; PID:g142786
A;Experimental source: strain OF4
C;Genetics:
C;Gene: ctaC
C;Superfamily: Cytochrome-c oxidase chain II, cytochrome c containing; cytochrome-c oxid
C;Keywords: copper; electron transfer; heme; heme-branched-associated complex; oxidoreductase
F;31-228/Domain: cytochrome-c oxidase chain II homolog <CO2>
F;175,210,214,221/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F;210,212,214,218/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F;212/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 25.6%; Score 52; DB 2; Length 342;
Best Local Similarity 33.3%; Pred. No. 6.1; Indels 10; Gaps 1;
Matches 13; Conservative 5; Mismatches 11;

Best Local Similarity 45.0%; Pred. No. 14;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 13 MLLARFRTRTPGDLHLDHSHV 32
Db 66 IILAKYRRKFGDDEIPKQVH 85

RESULT 10
E36792
hypothetical protein ORF58 - ictaluriid herpesvirus 1 (strain auburn 1)
C;Species: ictaluriid herpesvirus 1
A;Notes: host Ictalurus punctatus (channel catfish)
C;Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 09-Jul-2004
C;Accession: E36792
R;Davison, A.J.
submitted to GenBank, January 1992
A;Description: Channel catfish virus: a new type of herpesvirus.
A;Reference number: A36804
A;Accession: E36792
A;Molecule type: DNA
A;Residues: 1-599 <DAV>
A;Cross-references: UNIPROT:Q00157; GB:M75136; NID:G331209; PIDN:AAA88161.1; PID:G331268
R;Davison, A.J.
Virology 186, 9-14, 1992
A;Title: Channel catfish virus: a new type of herpesvirus.
A;Reference number: A39447; MUID:92087490; PMID:1727613
A;Contents: annotation
A;Note: neither protein nor nucleic acid sequence is given
C;Genetics:
A;Gene: 58
C;Superfamily: ictaluriid herpesvirus 1 hypothetical protein ORF58

Query Match 25.6%; Score 52; DB 2; Length 599;
Best Local Similarity 42.9%; Pred. No. 26;
Matches 12; Conservative 3; Mismatches 9; Indels 4; Gaps 1;

Qy 4 WIEMDTEMEMLLARFRTRTPGDLHLDHSHV 31
Db 260 WIKQDPAMAV---FRYPIDRLRLMI 263

RESULT 11
T51959
hypothetical protein [imported] - *Picea mariana*
C;Species: *Picea mariana*
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 09-Jul-2004
C;Accession: T51959
R;Perry, D.J.; Bousquet, J.
Genetics 149, 1089-1098, 1998
A;Title: Sequence-tagged-site (STS) markers of arbitrary genes. Development, characteriz
A;Reference number: Z25268; MUID:98278823; PMID:9611216
A;Accession: T51959
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-141 <PER>
A;Cross-references: UNIPROT:O65048; EMBL:AF051205; PIDN:AAC32110.1
A;Experimental source: embryo
C;Genetics:
A;Note: Sb08

Query Match 25.4%; Score 51.5; DB 2; Length 141;
Best Local Similarity 47.6%; Pred. No. 6.3; Indels 1; Gaps 1;
Matches 10; Conservative 7; Mismatches 3;

Qy 5 IEMDTEMEMLLARFRTRTPGDL 25
Db 65 LKMDIE-QQILSRQRPFGEL 84

RESULT 12
S26845
dTMP kinase (EC 2.7.4.9) - human

transcription factor isl-3 - chinook salmon
NAltName: names: insulin enhancer-binding protein isl-3; islet-3 protein
C.Species: Oncorhynchus tshawytscha (Chinook salmon)
C.Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 09-Jul-2004
C.Accession: C55973; S52090
R.Gong, Z.; Hui, C.; Hew, C.L.
J. Biol. Chem. 270, 3335-3345, 1995
A.Title: Presence of isl-1-related LIM domain homeobox genes in teleost and their simila
A.Reference number: A55973; MUID:95155429; PMID:7852419
A.Accession: C55973
A.Molecule type: mRNA
A.Residues: 1-363 <GON>
A.Cross-references: UNIPROT:P53409; GB:X64883

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 45.1796 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-33

Perfect score: 203

Sequence: 1 ALSWIEMDEMELARFRTPGDLHDSHVLCAP 37

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66.5	32.8	2156	2 Q7RQ36	Q7RQ36 plasmodium
2	66.5	32.8	2657	2 Q8ILG1	Q8ILG1 plasmodium
3	60	29.6	348	2 Q9K9N0	Q9K9N0 bacillus ha
4	56.5	27.8	1479	2 Q7FAB9	Q7FAB9 oryza sativ
5	56.5	27.8	1479	2 CAE03978	CAE03978 oryza sat
6	56.5	27.8	1708	2 Q7X2A3	Q7X2A3 oryza sativ
7	56	27.6	422	2 Q7MCN7	Q7MCN7 vibrio vuln
8	56	27.6	422	2 Q8D5N2	Q8D5N2 vibrio vuln
9	56	27.6	658	2 Q8DHF7	Q8DHF7 arabidopsis
10	55.5	27.3	428	2 Q9LH89	Q9LH89 arabidopsis
11	55.5	27.3	559	2 Q8RYX8	Q8RYX8 oryza sativ
12	55.5	27.3	1270	2 Q20749	Q20749 caenorhabdi
13	55	27.1	199	1 AC22_TENMO	F26368 tenebrio mo
14	54.5	26.8	213	2 Q8KUM3	Q8KUM3 erwinia chr
15	54.5	26.8	387	1 ISDF_ZYMMO	Q9trzi z ispd/isfp
16	54.5	26.8	610	2 Q9XTA0	Q9xta0 equus cabal
17	54	26.6	106	1 ILB1_CAEEL	Q09626 caenorhabdi
18	54	26.6	312	2 Q7PX17	Q7px17 anopheles g
19	54	26.6	464	2 Q9AJC6	Q9ajc6 rattus norv
20	54	26.6	464	2 AAH62007	AAH62007 rattus no
21	54	26.6	464	2 CAE83920	CAE83920 rattus no
22	54	26.6	771	2 Q89EX1	Q89ex1 bradyrhizob
23	54	26.6	1061	2 Q6MQS6	Q6mgs6 bdellovibri
24	54	26.6	1061	2 CAE78371	CAE78371 bdellovibri
25	53.5	26.4	557	2 Q6MQJ4	Q6mqj4 bdellovibri
26	53.5	26.4	557	2 CAE78453	CAE78453 bdellovibri
27	53	26.1	96	2 Q6N5Z5	Q6n5z5 rhodopsu
28	53	26.1	96	2 CAE28265	CAE28265 rhodopsu
29	53	26.1	96	2 Q9LI69	Q9li69 arabidopsis
30	52.5	25.9	659	2 Q72M95	Q72m95 leptospira
31	52.5	25.9	659	2 Q8EYt2	Q8eyt2 leptospira

ALIGNMENTS

RESULT 1

Q7RQ36 PRELIMINARY; PRT; 2156 AA.
AC Q7RQ36; AC Q7RQ36; TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PYO1267;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perte M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.B., Riedmiller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoai A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
ET "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL0100031; EAA20560.1; -;
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phosphatase.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR Hypothetical protein.
SQ SEQUENCE 2156 AA; 249387 MW; 47PBCC1CA2236384 CRC64;
Query Match 32.8%; Score 66.5; DB 2; Length 2156;
Best Local Similarity 35.4%; Pred. No. 4.4; 13; Indels 1; Gaps 1;
Matches 13; Conservative 6; Mismatches 382

RESULT 2

Q8ILG1 PRELIMINARY; PRT; 2657 AA.
ID Q8ILG1
AC Q8ILG1;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=PF14_0282;
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.S., Kene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Perte M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrell B.;
 RT "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum".
 RL Nature 419:458-511(2002).
 DR ENBL; AE014820; AAN36895.1; -;
 DR GO; GO:0003933; F:acid phosphatase activity; IEA.
 DR InterPro; IPR000560; HisAc phosphat.
 DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 2657 AA; 306805 MW; 4AA679A6D91877C2 CRC64;
 Query Match 32.8%; Score 56.5; DB 2; Length 2657;
 Best Local Similarity 39.4%; Pred. No. 5.5;
 Matches 13; Conservative 6; Mismatches 13; Indels 1; Gaps 1;
 QY 3 SWISDMTEMEMLARFRRTPGDLHDSVHLCA 35
 DB 358 NWYNIENEVLRKTRFQ-PDDLHCSHSELCS 389
 RESULT 3
 Q9K3N0 PRELIMINARY; PRT; 348 AA.
 ID Q9K3N0
 AC Q9K3N0
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE Cytochrome caa3 oxidase (Subunit II).
 GN Name=ctaC;
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=C-125;
 RC MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR ENBL; AP001516; BAB06334.1; -;
 DR PIR; G83976; G83976.
 DR HSP; P00092; 3C2C.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
 DR GO; GO:0005507; F:copper ion binding; IEA.
 DR GO; GO:0004129; F:cytochrome-c oxidase activity; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001505; Copper Cua.
 DR InterPro; IPR008972; Cupredoxin.
 DR InterPro; IPR000345; Cytochrome_B.
 DR InterPro; IPR009056; Cytochrome_c.
 DR InterPro; IPR003088; Cyt_C1.

DR InterPro; IPR002327; Cyt_C1AB.
 DR InterPro; IPR002429; Cyt_c_ox_2.
 DR Pfam; PF00116; COX2; 1.
 DR Pfam; PF02790; COX2_TM; 1.
 DR Pfam; PF00034; Cytochrome_C; 1.
 DR PRINTS; PR01166; CYCOXIDASEII.
 DR ProDom; PD000131; Copper Cua; 1.
 DR ProDom; PD000375; Cyt_C1AB; 1.
 DR ProSITE; PS00078; COX2; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
 KW Electron transport; Oxidoreductase; Respiratory chain; Transmembrane;
 SQ SEQUENCE 348 AA; 38585 MW; EA287CDIAC9AE323 CRC64;
 Query Match 29.6%; Score 60; DB 2; Length 348;
 Best Local Similarity 60.0%; Pred. No. 4.7;
 Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 QY 13 MLLARFRRTPGDLHDSVH 32
 DB 66 IILARFRKPGDDELPGQVH 85
 RESULT 4
 Q7FAB9 PRELIMINARY; PRT; 1479 AA.
 ID Q7FAB9
 AC Q7FAB9
 DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
 DE OSJNBa0033H08.2 protein.
 GN Name=OSJNBa0033H08.2;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12447439;
 RA Feng Q., Zhang Y., Hao P., Wang S., Pu G., Huang Y., Li Y., Zhu J.,
 RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Fang Y.,
 RA Wang Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
 RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
 RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
 RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
 RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
 RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
 RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
 RA Han B.;
 RT "Sequence and analysis of rice chromosome 4.";
 RL Nature 420:316-320(2002).
 DR ENBL; AL662942; CAE03978.3; -;
 DR InterPro; IPR001584; Rve.
 DR InterPro; IPR008975; Viral_cap_coat.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF00665; rve; 1.
 DR Pfam; PF00098; zf-CCHC; 1.
 DR PRINTS; PR00939; C2HCZNFINGER.
 DR SMART; SM00343; ZNF_C2HC; 1.
 DR PROSITE; PS50158; ZF_CCHC; 1.
 SQ SEQUENCE 1479 AA; 162541 MW; 8A8E584CA2A33DDF CRC64;
 Query Match 27.8%; Score 56.5; DB 2; Length 1479;
 Best Local Similarity 46.9%; Pred. No. 80;
 Matches 15; Conservative 2; Mismatches 10; Indels 5; Gaps 1;
 QY 10 EHEMLLARFRRTPG-----DLHDSVHLCAH 36
 DB 513 EQRLLARIPRTPGRLYMLDLNLRPVLAAH 544
 RESULT 5

```
CAE03978
ID CAE03978 PRELIMINARY; PRT; 1479 AA.
AC CAE03978;
DT 02-MAR-2004 (T-EMBLrel. 27, Created)
DT 02-MAR-2004 (T-EMBLrel. 27, Last sequence update)
DE OSJNB0033H08.2 protein.
OS OSJNB0033H08.2.
GN Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza; Oryza sativa.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA MDLINE=22337377; PubMed=12447439;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
DR EMBL; AL662942; CAE03978.3;
SQ SEQUENCE 1479 AA; 162541 MW; 8A8E584CA2A33DDF CRC64;

Query Match 27.8%; Score 56.5; DB 2; Length 1479;
Best Local Similarity 46.9%; Pred. No. 80;
Matches 15; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

OY 10 EMBLLARPRTPG-----DLHLDSVHLCAH 36
Db 513 EQRLLARIPRTGRLYMLDINLARPVCLAAH 544

RESULT 6
ID Q7XEA3 PRELIMINARY; PRT; 1708 AA.
AC Q7XEA3;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DE Putative gag-pol polyprotein.
GN ORFNames=OSJNB0073D04.30;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC The Rice Chromosome 10 Sequencing Consortium;
RA "In-depth view of structure, activity, and evolution of rice
RT chromosome 10.";
RL Science 300:1566-1569(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017096; AAP53887.1;
DR Gramene; Q7XEA3;
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR001878; Znf_CCHC.
```

```
DR Pfam; PF00651; BTB; 2.
DR Pfam; PF00651; rve; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00097; BTB; 1.
DR PROSITE; PS00158; ZF_CCHC; 1.
KW Polyprotein.
SQ SEQUENCE 1708 AA; 188208 MW; 0B0DE550A15A746E CRC64;

Query Match 27.8%; Score 56.5; DB 2; Length 1708;
Best Local Similarity 46.9%; Pred. No. 94;
Matches 15; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

OY 10 EMBLLARPRTPG-----DLHLDSVHLCAH 36
Db 584 EQRLLARIPRTGRLYMLDINLARPVCLAAH 615

RESULT 7
ID Q7MCN7 PRELIMINARY; PRT; 422 AA.
AC Q7MCN7;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Tyrosyl-tRNA synthetase.
GN Name=VVA1349;
OS Vibrio vulnificus (strain VJ016).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=196600;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14656365;
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
RT "Comparative genome analysis of Vibrio vulnificus, a marine
RT pathogen.";
RL Genome Res. 13:2577-2587(2003).
DR EMBL; AP005349; BAC97375.1;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0004831; F:tyrosine-tRNA ligase activity; IEA.
DR GO; GO:0006437; F:tyrosyl-tRNA aminoacylation; IEA.
DR InterPro; IPR002942; S4.
DR InterPro; IPR002305; tRNA-synt_1b.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002307; Tyr_tRNA-synt_1b.
DR Pfam; PF01479; S4; 1.
DR Pfam; PF00579; tRNA-synt_1b; 1.
DR PRINTS; PR01840; TRNASYNTHYR.
DR TIGRFAMs; TIGR00234; TyrS; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE; PS00889; S4; 1.
KW Aminoacyl-tRNA synthetase.
SQ SEQUENCE 422 AA; 47161 MW; 0AC47AC4DBA654F5 CRC64;

Query Match 27.6%; Score 56; DB 2; Length 422;
Best Local Similarity 35.9%; Pred. No. 22;
Matches 14; Conservative 5; Mismatches 10; Indels 10; Gaps 1;

OY 5 IEMTEMLLARPRR-----TPGDHLHDSVHL 33
Db 16 IAQTDLLEQLALFRQPTLYCYGDFDTAGSLHGLVPL 54

RESULT 8
ID Q8DSN2 PRELIMINARY; PRT; 422 AA.
AC Q8DSN2;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
```

```

RESULT 9
Q8DH7F PRELIMINARY; PRT; 658 AA.
ID Q8DH7F
AC Q8DH7;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Bicarboxylate transport system ATP-binding protein.
GN Name=cmpC; OrderedlocusNames=tlr2002;
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
[1]
RN RN
RP RP
RC STRAIN=BP-1.
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.:
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RT DNA Res. 9:123-130(2002).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AP005375; BAC09554.1; -.
DR HSSP; Q58206; IF30.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0015112; F:nitrate transporter activity; IEA.

```

RESULT 11					
Q8RYX8					
ID	Q8RYX8	PRELIMINARY;	PRT;	559 AA.	
AC	Q8RYX8;				
DT	01-JUN-2002	(TREMBlrel. 21, Created)			
DT	01-JUN-2002	(TREMBlrel. 21, Last sequence update)			

SQ SEQUENCE 199 AA; 20710 MW; 1EBF2D1DBEC6739F CRC64;
 Query Match 27.1%; Score 55; DB 1; Length 199;
 Best Local Similarity 52.9%; Pred. No. 13;
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 21 TPGDLHLHDHSHVHLCAHP 37
 Db 118 TPVDVHHEATHLKAHP 134

RESULT 14
 Q8KUM3 PRELIMINARY; PRT; 213 AA.
 AC Q8KUM3;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE HrpI.
 GN Name=hrpI;
 OS Erwinia chrysanthemi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Pectobacterium.
 OX NCBI_TaxID=5556;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ham J.H., Alfano J.R., Rodriguez-Palenzuela P., Rojas C.M.,
 RA Chatterjee A.K., Collmer A.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RA Rojas C.M., Ham J.H., Kim J.F., Beer S.V., Collmer A.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 CC -! SIMILARITY: Contains 1 HTH LuxR-type DNA-binding domain.
 CC EMBL; AF448202; AA046692.1; -.
 CC HSSP; P52934; 1QMP.
 CC DR GO; GO:0005622; C:intracellular; IEA.
 CC DR GO; GO:0003700; F:transcription factor activity; IEA.
 CC DR GO; GO:0001556; F:two-component response regulator activity; IEA.
 CC DR GO; GO:0007600; P:sensory perception; IEA.
 CC DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
 CC DR InterPro; IPR011006; CheY like.
 CC DR InterPro; IPR000792; HTH LuxR.
 CC DR InterPro; IPR001789; Response_reg.
 CC Pfam; PF00196; GerE; 1.
 CC Pfam; PF00072; Response_reg; 1.
 CC PRINTS; PR00038; HTHLUXR.
 CC ProDom; PD000307; HTH LuxR; 1.
 CC ProDom; PD000039; Response_reg; 1.
 CC SMART; SM00421; HTH LUXR; 1.
 CC SMART; SM00448; REC; 1.
 CC DR PROSITE; PS00622; HTH LUXR FAMILY; 1.
 CC DR PROSITE; PS01110; RESPONSE_REGULATORY; 1.
 KW DNA-binding; Phosphorylation; Sensory transduction;
 SQ Transcription regulation.
 SQ SEQUENCE 213 AA; 23772 MW; 728465D54BEB4249 CRC64;
 Query Match 26.8%; Score 54.5; DB 2; Length 213;
 Best Local Similarity 40.6%; Pred. No. 17;
 Matches 13; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

OY 1 ALSWIENDTMEMLARFR-TPGDHLHDHSHV 31
 Db 101 ALGYITKDPEALLSRIREVAQGYIDHSI 132

RESULT 15
 IDSF_ZYMMO STANDARD; PRT; 387 AA.
 AC Q9RNL; Q9S1D5; Q9S1D6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE IspD/ispF bifunctional enzyme [Includes: 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60) (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP cytidylyltransferase) (MCT); 2-C-methyl-D-erythritol 2,4-cyclophosphate synthase (EC 4.6.1.12) (MECPs) (MECDP-synthase)].
 GN Name=ispDF;
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
 CC Sphingomonadaceae; Zymomonas.
 OX NCBI_TaxID=542;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 31821 / ZM4 / CP4;
 RA Lee H.J., Kang H.S.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 31821 / ZM4 / CP4;
 RX MEDLINE=99350425; PubMed=10419959;
 RA Douka E., Koukhou A.I., Vatholomatos G., Frilligos S.,
 RA Papamichael E.M., Drinas C.;
 RT "A Zymomonas mobilis mutant with delayed growth on high glucose concentrations";
 RL J. Bacteriol. 181:4598-4604(1999).
 CC -! FUNCTION: Bifunctional enzyme that catalyzes the formation of 4-diphosphocytidylyl-2C-methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-phosphate (ispD), and converts 4-diphosphocytidylyl-2C-methyl-D-erythritol 2-phosphate into 2C-methyl-D-erythritol 2,4-cyclophosphate and CMP (ispF). Also converts 4-diphosphocytidylyl-2C-methyl-D-erythritol into 2C-methyl-D-erythritol 3,4-cyclophosphate and CMP (ispF) (By similarity).
 CC -! CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate = diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -! CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclophosphate + CMP.
 CC -! PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC -! PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fifth step.
 CC -! SIMILARITY: In the N-terminal section; belongs to the ispD family.
 CC -! SIMILARITY: In the C-terminal section; belongs to the ispF family.
 CC -! CAUTION: Ref.2 sequence differs from that shown due to frameshifts.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 EMBL; AF176314; AAD53897.1; -.
 EMBL; AJ009974; CAB51894.1; ALT_FRAME.
 EMBL; AJ009974; CAB51895.1; ALT_FRAME.
 HSSP; Q8RQF5; 11V1.
 HAMAP; MF_00107; fused; 1.
 HAMAP; MF_00108; fused; 1.
 InterPro; IPR008234; IspD/IspF synth.
 InterPro; IPR001228; IspD synthase.
 InterPro; IPR003526; YgbB.
 Pfam; PF01128; IspD; 1.
 Pfam; PF02542; YgbB; 1.
 TIGRFAMs; TIGR00151; ispF; 1.
 PROSITE; PS01295; IspD; 1.
 PROSITE; PS01350; IspF; 1.
 Isoprene biosynthesis; Lyase; Multifunctional enzyme;
 Nucleotidyltransferase; transferase.
 DOMAIN 1 223
 4-diphosphocytidylyl-2C-methyl-D-erythritol synthase.
 DOMAIN 224 387
 2-C-methyl-D-erythritol 2,4-cyclophosphate synthase.
 CONFLICT 30 31
 KT -> QI (in Ref. 2).

```

FT CONFLICT 36      36      N -> I (in Ref. 2).
FT CONFLICT 63      63      T -> A (in Ref. 2).
SQ SEQUENCE 387 AA; 41946 MW; A99580SEFD952B27 CRC64;

Query Match      26.8%; Score 54.5; DB 1; Length 387;
Best Local Similarity 30.3%; Pred.No. 34;
Matches 10; Conservative 13; Mismatches 9; Indels 1; Gaps 1;

QY      6 EMDTEMEMLLARFRPTPGDLHLHDSV-HLCAHP 37
      : : : : : : : : : : : : : : : : : :
Db      16 QVNVQVKALQKYRKTAGKANLAHAIDNLLAHP 48

```

Search completed: November 10, 2004, 13:38:33
Job time : 47.1796 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 3.79965 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-34
Perfect score: 57
Sequence: 1 TSTLPHRRR 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB pep.*
3: /cgn2_6/prodata/1/iaa/6A COMB pep.*
4: /cgn2_6/prodata/1/iaa/6B COMB pep.*
5: /cgn2_6/prodata/1/iaa/6CTUS COMB pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1 pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	36	63.2	54	4	US-09-621-976-5801
2	36	63.2	265	4	US-09-270-767-45056
3	36	63.2	381	4	US-09-510-031A-5
4	36	63.2	420	4	US-09-583-110-4701
5	36	63.2	449	2	US-08-819-458A-16
6	35	61.4	325	4	US-09-252-991A-22955
7	35	61.4	362	4	US-09-393-627B-3
8	35	61.4	362	4	US-09-758-008-2
9	35	61.4	362	4	US-09-758-008-4
10	35	61.4	638	2	US-08-681-151-3
11	35	61.4	884	2	US-08-465-376A-2
12	35	61.4	884	2	US-08-982-412-2
13	34	59.6	144	4	US-10-101-464A-555
14	34	59.6	146	4	US-09-252-991A-21166
15	34	59.6	159	4	US-09-107-532A-4074
16	34	59.6	252	4	US-09-270-767-40047
17	34	59.6	252	4	US-09-270-767-55263
18	34	59.6	510	4	US-09-489-039A-12958
19	34	59.6	560	4	US-09-489-039A-11540
20	34	59.6	853	2	US-08-468-558-3
21	34	59.6	853	3	US-08-676-444-3
22	34	59.6	1226	4	US-09-601-537-7
23	34	59.6	1227	2	US-08-760-075A-18
24	34	59.6	1227	3	US-09-338-546-18
25	34	59.6	1227	4	US-09-659-084-18
26	33	57.9	28	4	US-09-270-767-57530
27	33	57.9	153	4	US-09-270-767-45940

28 57.9 197 4 US-09-252-991A-31560 Sequence 31560, A
29 57.9 231 1 US-08-144-121-5 Sequence 5, Appli
30 57.9 231 2 US-08-735-893-5 Sequence 5, Appli
31 57.9 233 4 US-09-270-767-42251 Sequence 42251, A
32 57.9 248 2 US-08-867-087B-17 Sequence 17, Appl
33 57.9 249 2 US-08-867-087B-15 Sequence 15, Appl
34 57.9 314 4 US-09-543-681A-7410 Sequence 7410, Ap
35 57.9 314 1 US-07-930-686-12 Sequence 12, Appl
36 57.9 440 2 US-08-460-398-12 Sequence 12, Appl
37 57.9 440 2 US-09-710-279-1832 Sequence 1832, Ap
38 57.9 493 4 US-09-252-991A-27294 Sequence 27294, A
39 57.9 509 3 US-09-199-637A-321 Sequence 321, App
40 57.9 529 4 US-09-252-991A-21425 Sequence 21425, A
41 57.9 531 4 US-09-252-991A-24404 Sequence 24404, A
42 57.9 593 3 US-09-134-001C-3991 Sequence 3991, Ap
43 57.9 825 4 US-09-270-767-45321 Sequence 45321, A
44 57.9 1147 1 US-08-144-121-3 Sequence 3, Appli
45 57.9 1147 2 US-08-735-893-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-621-976-5801
; Sequence 5801, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 5801
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -46..-1
US-09-621-976-5801

Query Match 63.2%; Score 36; DB 4; Length 54;
Best Local Similarity 87.5%; Pred. No. 7.6;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 STLPHRR 9
Db 43 STLPHSR 50

RESULT 2
US-09-270-767-45056
; Sequence 45056, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45056
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid

US-09-270-767-45056

Query Match 63.2%; Score 36; DB 4; Length 265;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPPIRR 9
DB 124 TQALPHLR 132

RESULT 3

US-09-510-031A-5
; Sequence 5, Application US/09510031A
; Patent No. 6638518
; GENERAL INFORMATION:
; APPLICANT: Ratliff, Timothy
; APPLICANT: Kline, Joel
; TITLE OF INVENTION: METHOD FOR INHIBITING INFLAMMATORY RESPONSES
; FILE REFERENCE: 140.0010 0101
; CURRENT APPLICATION NUMBER: US/09/510,031A
; CURRENT FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/121,177
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Mycobacterium avium
US-09-510-031A-5

Query Match 63.2%; Score 36; DB 4; Length 381;
Best Local Similarity 63.6%; Pred. No. 63;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSTLPPIRRTR 11
DB 51 TATTPHRRR 61

RESULT 4

US-09-583-110-4701
; Sequence 4701, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4701
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae

QY 3 TLPHRR 10
DB 340 TLPHLR 347

Query Match 63.2%; Score 36; DB 4; Length 420;
Best Local Similarity 75.0%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 5

US-08-819-458A-16
; Sequence 16, Application US/08819458A
; Patent No. 5891669
; GENERAL INFORMATION:
; APPLICANT: Jensen, Ejner B.
; APPLICANT: Cherry, Joel
; APPLICANT: Elrod, Susan L.
; TITLE OF INVENTION: Methods For Producing Polypeptides
; TITLE OF INVENTION: In Respiratory-Deficient Cells
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58916690 No. 5891669disk of No. 5891669th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/819,458A
FILING DATE: 17-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5215.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-819-458A-16

Query Match 63.2%; Score 36; DB 2; Length 449;
Best Local Similarity 72.7%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSTLPPIRRTR 11
DB 195 TSMPLKIRSTR 205

RESULT 6

US-09-252-991A-22955
; Sequence 22955, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22955
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-22955

Query Match 61.4%; Score 35; DB 4; Length 325;
Best Local Similarity 63.6%; Pred. No. 82;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 11
Db 273 SSTLPALRRR 283

RESULT 7

US-09-393-627B-3

; Sequence 3, Application US/09393627B
; Patent No. 6455314
; GENERAL INFORMATION:
; APPLICANT: Wickham, Thomas J.
; APPLICANT: Kovesdi, Imre
; APPLICANT: Roelvink, Petrus W.
; APPLICANT: Bruder, Joseph T.
; TITLE OF INVENTION: Alternatively Targeted Adenovirus
; FILE REFERENCE: 202345
; CURRENT APPLICATION NUMBER: US/09/393,627B
; CURRENT FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 60/099,851
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: US 60/136,529
; PRIOR FILING DATE: 1999-05-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.2
; SEQ ID NO 3
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Human adenovirus serotype 9

US-09-393-627B-3

Query Match 61.4%; Score 35; DB 4; Length 362;
Best Local Similarity 70.0%; Pred. No. 92;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
Db 123 TSTLPLRLNT 132

RESULT 8

US-09-758-008-2

; Sequence 2, Application US/09758008
; Patent No. 6635466
; GENERAL INFORMATION:
; APPLICANT: Law, Lane K.
; APPLICANT: Davidson, Beverly L.
; TITLE OF INVENTION: Adenovirus serotype 30 (Ad30)
; FILE REFERENCE: 875.044US1
; CURRENT APPLICATION NUMBER: US/09/758,008
; CURRENT FILING DATE: 2001-01-09
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Adenovirus

US-09-758-008-2

Query Match 61.4%; Score 35; DB 4; Length 362;
Best Local Similarity 70.0%; Pred. No. 82;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
Db 123 TSTLPLRLNT 132

RESULT 9

US-09-758-008-4

; Sequence 4, Application US/09758008
; Patent No. 6635466
; GENERAL INFORMATION:
; APPLICANT: Law, Lane K.
; APPLICANT: Davidson, Beverly L.
; TITLE OF INVENTION: Adenovirus serotype 30 (Ad30)
; FILE REFERENCE: 875.044US1
; CURRENT APPLICATION NUMBER: US/09/758,008
; CURRENT FILING DATE: 2001-01-09
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Adenovirus

US-09-758-008-4

Query Match 61.4%; Score 35; DB 4; Length 362;
Best Local Similarity 70.0%; Pred. No. 92;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
Db 123 TSTLPLRLNT 132

RESULT 10

US-08-681-151-3

; Sequence 3, Application US/08681151
; Patent No. 5869637
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Braxton, Scott Michael
; APPLICANT: Goli, Surya
; TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,151
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0074US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

QY 1 TSTLPHIRRT 10
Db 123 TSTLPLRLNT 132

```

; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 205011
US-08-681-151-3

Query Match 61.4%; Score 35; DB 2; Length 638;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 STLPHIRTR 10
DB 84 TGTLPRIHRT 93

RESULT 11
US-08-465-976A-2
; Sequence 2, Application US/08465976A
; Patent No. 5869632
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN GILFILLAN, CECCHI
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: US
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,976A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY F
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-444
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 994-1700
; TELEFAX: (201) 994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-976A-2

Query Match 61.4%; Score 35; DB 2; Length 884;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 STLPHIRTR 11
DB 716 STLPPPRTR 725

RESULT 12
US-08-982-412-2
; Sequence 2, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 205011
US-08-681-151-3

Query Match 61.4%; Score 35; DB 2; Length 638;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 STLPHIRTR 10
DB 84 TGTLPRIHRT 93

RESULT 11
US-08-465-976A-2
; Sequence 2, Application US/08465976A
; Patent No. 5869632
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN GILFILLAN, CECCHI
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: US
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,976A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY F
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-444
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 994-1700
; TELEFAX: (201) 994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-976A-2

Query Match 61.4%; Score 35; DB 2; Length 884;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 STLPHIRTR 11
DB 716 STLPPPRTR 725

RESULT 12
US-08-982-412-2
; Sequence 2, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 205011
US-08-681-151-3

Query Match 61.4%; Score 35; DB 2; Length 638;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 STLPHIRTR 10
DB 84 TGTLPRIHRT 93

RESULT 11
US-08-465-976A-2
; Sequence 2, Application US/08465976A
; Patent No. 5869632
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN GILFILLAN, CECCHI
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: US
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,976A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY F
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-444
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 994-1700
; TELEFAX: (201) 994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-976A-2

Query Match 61.4%; Score 35; DB 2; Length 884;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 STLPHIRTR 11
DB 716 STLPPPRTR 725

RESULT 12
US-08-982-412-2
; Sequence 2, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 205011
US-08-681-151-3

Query Match 61.4%; Score 35; DB 2; Length 638;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 STLPHIRTR 10
DB 84 TGTLPRIHRT 93

RESULT 11
US-08-465-976A-2
; Sequence 2, Application US/08465976A
; Patent No. 5869632
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN GILFILLAN, CECCHI
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: US
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,976A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY F
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-444
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 994-1700
; TELEFAX: (201) 994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-976A-2

Query Match 61.4%; Score 35; DB 2; Length 884;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 STLPHIRTR 11
DB 716 STLPPPRTR 725

RESULT 12
US-08-982-412-2
; Sequence 2, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 205011
US-08-681-151-3

Query Match 61.4%; Score 35; DB 2; Length 638;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 STLPHIRTR 10
DB 84 TGTLPRIHRT 93

RESULT 11
US-08-465-976A-2
; Sequence 2, Application US/08465976A
; Patent No. 5869632
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN GILFILLAN, CECCHI
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: US
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,976A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY F
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-444
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 994-1700
; TELEFAX: (201) 994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-976A-2

Query Match 61.4%; Score 35; DB 2; Length 884;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 STLPHIRTR 11
DB 716 STLPPPRTR 725

RESULT 12
US-08-982-412-2
; Sequence 2, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 205011
US-08-681-151-3

Query Match 61.4%; Score 35; DB 2; Length 638;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 STLPHIRTR 10
DB 84 TGTLPRIHRT 93

RESULT 11
US-0
```

```
Query Match      59.6%; Score 34; DB 4; Length 144;
Best Local Similarity 70.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TSTLPHIRRT 10
Db 26 TSLNPHIRNT 35

RESULT 14
US-09-252-991A-21166
; Sequence 21166, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21166
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21166

Query Match      59.6%; Score 34; DB 4; Length 146;
Best Local Similarity 77.8%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TLPHIRTR 11
Db 72 TTPHQRRTR 80

RESULT 15
US-09-107-532A-4074
; Sequence 4074, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4074:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...159
; SEQUENCE DESCRIPTION: SEQ ID NO: 4074:
US-09-107-532A-4074

Query Match      59.6%; Score 34; DB 4; Length 159;
Best Local Similarity 63.6%; Pred. No. 58;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TSTLPHIRTR 11
Db 53 SSSLHLTRTR 63

Search completed: November 10, 2004, 13:44:06
Job time : 4.79965 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 11.8169 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-34
Perfect score: 57
Sequence: 1 TSTLPHIRTR 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata1/pubpaa/PCT05_PUBCOMB.pep.*
7: /cgn2_6/ptodata1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	100.0	11	14	US-10-092-750-34
2	49	86.0	28	15	Sequence 34, Appl
3	42	73.7	460	16	Sequence 242274,
4	40	70.2	339	14	Sequence 202505,
5	39	68.4	96	15	Sequence 8110, Ap
6	38	66.7	131	15	Sequence 252181,
7	38	66.7	176	16	Sequence 57404, A
8	38	66.7	898	15	Sequence 46417, A
9	38	66.7	903	14	Sequence 77110, A
10	38	66.7	1224	14	Sequence 4205, Ap
11	38	66.7	1234	14	Sequence 23379, A
12	37	64.9	170	11	Sequence 7454, Ap
13	37	64.9	100	16	Sequence 6100, Ap
					Sequence 189053,

14	64.9	116	15	US-10-424-599-242031	Sequence 242031,
15	64.9	208	16	US-10-437-963-170498	Sequence 170498,
16	64.9	223	17	US-10-425-115-255867	Sequence 255867,
17	64.9	272	14	US-10-369-493-10497	Sequence 10497, A
18	64.9	310	9	US-09-738-626-4591	Sequence 4591, Ap
19	64.9	598	16	US-10-437-963-180630	Sequence 180630,
20	64.9	646	16	US-10-437-963-152417	Sequence 152417,
21	64.9	887	14	US-10-369-493-19640	Sequence 19640, A
22	64.9	1026	14	US-10-369-493-12400	Sequence 12400, A
23	64.9	1209	14	US-10-128-714-3144	Sequence 3144, Ap
24	64.9	1255	14	US-10-128-714-8144	Sequence 8144, Ap
25	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
26	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
27	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
28	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
29	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
30	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
31	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
32	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
33	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
34	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
35	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
36	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
37	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
38	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
39	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
40	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
41	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
42	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
43	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
44	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap
45	64.9	1261	15	US-10-320-797-3345	Sequence 3345, Ap

ALIGNMENTS

RESULT 1
US-10-092-750-34
; Sequence 34, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-34

Query Match 100.0%; Score 57; DB 14; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSTLPHIRTR 11
DB 1 TSTLPHIRTR 11
RESULT 2
US-10-424-599-242274
; Sequence 242274, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 242274
 ; LENGTH: 28
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_60801C.1.pep
 US-10-424-599-242274

Query Match 86.0%; Score 49; DB 15; Length 28;
 Best Local Similarity 90.0%; Pred. No. 0.13; Mismatches 1; Indels 0; Gaps 0;
 Matches 9; Conservative 0;

QY 2 STLPHIRRTTR 11
 :|||||
 DB 15 ATLPHIRRTTR 24

RESULT 3
 US-10-437-963-202505
 ; Sequence 202505, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 202505
 ; LENGTH: 460
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_97779C.1.pep
 US-10-437-963-202505

Query Match 73.7%; Score 42; DB 16; Length 460;
 Best Local Similarity 72.7%; Pred. No. 40; Mismatches 2; Indels 1; Gaps 0;
 Matches 8; Conservative 1;

QY 1 TSTLPHIRRTTR 11
 :|||||
 DB 442 TSTVFNIRTR 452

RESULT 4
 US-10-156-761-8110
 ; Sequence 8110, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 8110
 ; LENGTH: 339
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-8110

Query Match 70.2%; Score 40; DB 14; Length 339;
 Best Local Similarity 70.0%; Pred. No. 67; Mismatches 1; Indels 2; Gaps 0;
 Matches 7; Conservative 1;

QY 2 STLPHIRRTTR 11
 :|||||
 DB 275 STPLHRRAR 284

RESULT 5
 US-10-424-599-252181
 ; Sequence 252181, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 252181
 ; LENGTH: 96
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_69748C.1.pep
 US-10-424-599-252181

Query Match 68.4%; Score 39; DB 15; Length 96;
 Best Local Similarity 63.6%; Pred. No. 28; Mismatches 2; Indels 0; Gaps 0;
 Matches 7; Conservative 2;

QY 1 TSTLPHIRRTTR 11
 :|||||
 DB 64 TTTLPHLPSTR 74

RESULT 6
 US-10-425-114-57404
 ; Sequence 57404, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E.
 ; APPLICANT: Tabaska, Jack E.
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114

```
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57404
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17344G01_FLI.pep
US-10-425-114-57404

Query Match 66.7%; Score 38; DB 15; Length 131;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
   |||||
Db 1 TSSLPAIRT 10

RESULT 7
US-10-767-701-46417
; Sequence 46417, Application US/10767701
; Publication No. US20040172694A1
; GENERAL INFORMATION:
; APPLICANT: Kowalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 46417
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C29621_1.pep
US-10-767-701-46417

Query Match 66.7%; Score 38; DB 16; Length 176;
Best Local Similarity 75.0%; Pred. No. 78;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 STLPHIR 9
   |||||
Db 157 ATLPHVR 164

RESULT 8
US-10-282-122A-77110
; Sequence 77110, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chisen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/131,078
; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 77110
; LENGTH: 898
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-77110

Query Match 66.7%; Score 38; DB 15; Length 898;
Best Local Similarity 80.0%; Pred. No. 4.1e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
   |||||
Db 413 TSTLDYIRRT 422

RESULT 9
US-10-369-493-4205
; Sequence 4205, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4205
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-4205

Query Match 66.7%; Score 38; DB 14; Length 903;
Best Local Similarity 70.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
   |||||
Db 287 SSTLSHLART 296

RESULT 10
US-10-369-493-22379
; Sequence 22379, Application US/10369493
; Publication No. US20030233675A1
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6100
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-6100

Query Match          64.9%; Score 37; DB 11; Length 70;
Best Local Similarity 75.0%; Pred. No. 46;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 STLPHIRRT 9
Db 30 STLPHLRK 37

RESULT 13
US-10-437-963-189053
; Sequence 189053, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 189053
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85599C.1.pep
US-10-437-963-189053

Query Match          64.9%; Score 37; DB 16; Length 100;
Best Local Similarity 70.0%; Pred. No. 66;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
Db 69 TSLGPLHAKRT 78

RESULT 14
US-10-424-599-242031
; Sequence 242031, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 242031
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure

; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22379
; LENGTH: 1224
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22379

Query Match          66.7%; Score 38; DB 14; Length 1224;
Best Local Similarity 70.0%; Pred. No. 5.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
Db 489 SSTLSHLRRT 498

RESULT 11
US-10-032-585-7454
; Sequence 7454, Application US/10032585
; Publication No. US20030180553A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-085-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7454
; LENGTH: 1234
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7454

Query Match          66.7%; Score 38; DB 14; Length 1234;
Best Local Similarity 70.0%; Pred. No. 5.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
Db 484 SSTLSHLRRT 493

RESULT 12
US-09-864-408A-6100
; Sequence 6100, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
```

```

; LOCATION: (1)..(116)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60581C.1.pap
US-10-424-599-242031
Query Match      64.9%; Score 37; DB 15; Length 116;
Best Local Similarity 87.5%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3  TLPHIRRT 10
Db      104 TLPHIRMT 111

RESULT 15
US-10-437-963-170498
; Sequence 170498, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 170498
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(208)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_69819C.1.pap
US-10-437-963-170498
Query Match      64.9%; Score 37; DB 16; Length 208;
Best Local Similarity 72.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  TSTLPHIRTR 11
Db      15 TSTTPRRTR 25

Search completed: November 11, 2004, 01:28:16
Job time : 12.8669 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 2.45078 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-34
Perfect score: 57
Sequence: 1 TSTLPHRRTR 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	70.2	868	2 F89500	protein T27A10.6 [
2	38	66.7	42	2 F82691	hypothetical prote
3	38	66.7	738	2 S01892	hemolysin A precu
4	38	66.7	898	2 A82298	translational initia
5	38	66.7	1224	2 A25884	DNA-directed RNA p
6	37	64.9	383	2 D89975	conserved hypothet
7	37	64.9	1210	2 S35548	DNA-directed RNA p
8	37	64.9	1879	2 T15962	hypothetical prote
9	36	63.2	64	2 C64010	hypothetical prote
10	36	63.2	107	2 T43914	probable integral
11	36	63.2	314	2 T36913	gamma-1-microglobu
12	36	63.2	355	1 S22181	IS1167, transposas
13	36	63.2	422	2 C95053	hypothetical prote
14	36	63.2	772	2 T30645	UL47h protein - Ca
15	36	63.2	808	2 JQ2205	unc-5 protein - Ca
16	36	63.2	919	2 T32541	unc-5 protein, lon
17	36	63.2	947	1 B44294	conserved hypothet
18	36	63.2	1095	2 AD0301	chaoptin precursor
19	36	63.2	1134	1 A29944	conserved hypothet
20	35	61.4	278	2 E75546	insulin-like prote
21	35	61.4	305	2 A38422	fiber protein - hu
22	35	61.4	362	2 S37220	fiber protein - hu
23	35	61.4	362	2 S40092	hypothetical prote
24	35	61.4	366	2 T25178	probable cycA prot
25	35	61.4	556	2 T70503	beta-glucosidase h
26	35	61.4	591	2 T02402	plasma kallikrein
27	35	61.4	638	1 KQMSPL	plasma kallikrein
28	35	61.4	638	1 KQRTFL	glutamate synthase
29	35	61.4	1137	2 T02097	

ALIGNMENTS

RESULT 1

F89500
protein T27A10.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: F89500
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology;
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: F89500
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-868 <STO>
A:Cross-references: UNIPROT:Q22822; GB:chr_X; PIDN:AAB52504.1; PID:g1326289; GSPDB:GN000;
C:Genetics:
A:Gene: T27A10.6
A:Map position: X

Query Match 70.2%; Score 40; DB 2; Length 868;
Best Local Similarity 72.7%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSTLPHRRTR 11
DB 556 TKTTPRRRTR 566

RESULT 2

F82691
hypothetical protein XF1351 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: F82691
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20165717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: F82691
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-42 <SIM>
A:Cross-references: UNIPROT:Q9PDM8; GB:AE003967; GB:AE003849; NID:g9106347; PIDN:AAF84161
R:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme

MEGF2 protein - hu
ferredoxin-depende
glutaredoxin (grx-
probable acetyltra
hypothetical prote
hypothetical prote
conserved hypothet
probable lipu prot
hypothetical prote
amino acid regulat
hypothetical prote
hypothetical prote
translation initia
probable ATP-depen
hypothetical prote
coagulation factor

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigh Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Merck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1351

Query Match 66.7%; Score 38; DB 2; Length 42;
Best Local Similarity 70.0%; Pred. No. 1.8;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

Qy 2 STLPPIRRTR 11
|:|:|:|:|
Db 17 STMPHQHTR 26

RESULT 3
hemolysin A precursor - Vibrio cholerae (strain El Tor 017)
N;Alternate names: El Tor hemolysin
C;Species: Vibrio cholerae
C;Date: 30-Jun-1989 #sequence_revision 10-Jun-1989 #text_change 10-Dec-1999
C;Accession: S01892
R;Alm, R.A.; Strocher, U.H.; Manning, P.A.
Mol. Microbiol. 2, 481-488, 1988
A;Title: Extracellular proteins of Vibrio cholerae: nucleotide sequence of the structural in the non-haemolytic classical strain 569B.
A;Reference number: S01892; MUID:89013889; PMID:30503359
A;Accession: S01892
A;Molecule type: DNA
A;Residues: 1-738 <ALM>
A;Cross-references: EMBL:Y00557
C;Genetics:
A;Gene: hlyA
C;Superfamily: Vibrio hemolysin A
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-738/Product: hemolysin A #status predicted <MAT>

Query Match 66.7%; Score 38; DB 2; Length 738;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 2; Mismatches 0; Gaps 0;

Qy 1 TSTLPPIR 8
|:|:|:|
Db 159 TNLPHVR 166

RESULT 4
A82298
translation initiation factor IF-2 VC0643 [imported] - Vibrio cholerae (strain N16961 se
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: A82298
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bacs, S.; Qin, H.; Dragoi, I.; Sellers, F. l., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: A82298
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-898 <HEI>
A;Cross-references: UNIPROT:Q9KU80; GB:AE004150; GB:AE003852; NID:g9655075; PIDN:AAF9380
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0643
A;Map position: 1

C;Superfamily: translation initiation factor IF-2; translation elongation factor Tu homolog
F;400-509/Domain: translation elongation factor Tu homology <ETU>

Query Match 66.7%; Score 38; DB 2; Length 898;
Best Local Similarity 80.0%; Pred. No. 42;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TSTLPPIRRT 10
|:|:|:|:|
Db 413 TSTLDYIRRT 422

RESULT 5
A25884
DNA-directed RNA polymerase (EC 2.7.7.6) II 140K chain [validated] - Yeast (Saccharomyces
N;Alternate names: DNA-directed RNA polymerase B chain B150; protein Q3533; protein YOR1
C;Species: Saccharomyces cerevisiae
C;Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 09-Jul-2004
C;Accession: A25884; S36860; A37894; S67039
R;Sweetser, D.; Nonet, M.; Young, R.A.
Proc. Natl. Acad. Sci. U.S.A. 84, 1192-1196, 1987
A;Title: Prokaryotic and eukaryotic RNA polymerases have homologous core subunits.
A;Reference number: A25884; MUID:87147239; PMID:3547406
A;Accession: A25884
A;Molecule type: DNA
A;Residues: 1-1224 <SWE>
A;Cross-references: UNIPROT:P08518; EMBL:M15693; NID:g172210; PIDN:AAA68096.1; PID:g17221
R;Sweetser, D.; Nonet, M.; Young, R.A.
submitted to the EMBL Data Library, July 1987
A;Description: Prokaryotic and eukaryotic RNA polymerases have homologous core subunits.
A;Reference number: S36860
A;Accession: S36860
A;Molecule type: DNA
A;Residues: 1-1002, 'RRRY', 1007-1224 <SW2>
A;Cross-references: EMBL:M15693; NID:g172210; PIDN:AAA68096.1; PID:g172211
R;Riva, M.; Carles, C.; Sentenac, A.; Grachev, M.A.; Mustaev, A.A.; Zaychikov, E.F.
J. Biol. Chem. 265, 16498-16503, 1990
A;Title: Mapping the active site of yeast RNA polymerase B (II).
A;Reference number: A37894; MUID:90375520; PMID:2204624
A;Accession: A37894
A;Molecule type: protein
A;Residues: 961-990 <RIV>
R;Bordone, R.; Camasses, A.; Madania, A.; Martin, R.P.; Poch, O.; Tarassov, I.A.; Winsor
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67032
A;Accession: S67039
A;Molecule type: DNA
A;Residues: 1-1224 <BOR>
A;Cross-references: EMBL:Z75059; NID:g1420378; PIDN:CAA99357.1; PID:g1420379; MIPS:YOR151
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:RPB2
A;Cross-references: SGD:S0005677; MIPS:YOR151c
A;Map position: 15R
C;Function:
A;Description: EC 2.7.7.6 [validated, MUID:87147239]
C;Superfamily: DNA-directed RNA polymerase 132K polypeptide
C;Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription; zinc finger

Query Match 66.7%; Score 38; DB 2; Length 1224;
Best Local Similarity 70.0%; Pred. No. 58;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TSTLPPIRRT 10
|:|:|:|:|
Db 489 SSTLSHIRT 498

RESULT 6
DB9975
conserved hypothetical protein SA1693 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C;Accession: D89975
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C;Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: D89975
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-383 <KUR>
A;Cross-references: UNIPROT:Q99T04; GB:BA000018; PID:G13701670; PIDN:BA042963.1; GSPDB:G
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA1693

Query Match 64.9%; Score 37; DB 2; Length 383;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 STLPHRR 9
Db 195 STLPHLRK 202
|||||

RESULT 7
S35548
DNA-directed RNA polymerase (EC 2.7.7.6) II chain 2 - fission yeast (*Schizosaccharomyces*
C;Species: *Schizosaccharomyces pombe*
C;Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S35548
R;Kawagishi, M.; Yamagishi, M.; Ishihama, A.
Nucleic Acids Res. 21, 469-473, 1993
A;Title: Cloning and sequence determination of the *Schizosaccharomyces pombe* rpb2 gene
A;Reference number: S35548; MUID:93181236; PMID:8441660
A;Accession: S35548
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1210 <KAW>
A;Cross-references: UNIPROT:Q02061; EMBL:D13337; NID:G218553; PIDN:BAA02600.1; PID:G2185
A;Note: part of this sequence was confirmed by mRNA sequencing
R;Kawagishi-Kobayashi, M.; Yamamoto, M.; Ishihama, A.
Mol. Gen. Genet. 250, 1-6, 1996
A;Title: Mutational analysis of the RNase-like domain in subunit 2 of fission yeast RNA
A;Reference number: S61777; MUID:96158839; PMID:8569679
A;Contents: annotation; mutational analysis
C;Genetics:
A;Gene: rpb2
A;Introns: 6/1
A;Superfamily: DNA-directed RNA polymerase 132K polypeptide
C;Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription

Query Match 64.9%; Score 37; DB 2; Length 1210;
Best Local Similarity 77.8%; Pred. No. 89;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 STLPHRR 10
Db 476 STLPHRR 484
|||||

RESULT 8
T15962
hypotheical protein F07C7.1 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T15962
R;Favell, T.
submitted to the EMBL Data Library, January 1996
A;Description: The sequence of *C. elegans* cosmid F07C7.
A;Reference number: Z18437
A;Accession: T15962
A;Status: preliminary; translated from GB/EMBL/DD50

A;Molecule type: DNA
A;Residues: 1-1879 <FAY>
A;Cross-references: UNIPROT:Q19161; EMBL:U46672; NID:gl166599; PID:gl166600; PIDN:AAA857;
C;Genetics:
A;Gene: CESP:F07C7.1
A;Introns: 818/1; 1522/1; 1702/3; 1784/1

Query Match 64.9%; Score 37; DB 2; Length 1879;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LPHIRRT 11
Db 605 LPHIMRT 612
|||||

RESULT 9
C64010
hypotheical protein HI0592 - *Haemophilus influenzae* (strain Rd KW20)
C;Species: *Haemophilus influenzae*
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C;Accession: C64010
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: C64010
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-64 <TIGR>
A;Cross-references: UNIPROT:P44021; GB:U32741; GB:L42023; NID:gi573582; PIDN:AAC22259.1;

Query Match 63.2%; Score 36; DB 2; Length 64;
Best Local Similarity 70.0%; Pred. No. 6.5;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STLPHIRRT 11
Db 28 SLMAHRR 37
|:|||||

RESULT 10
T44914
hypotheical protein MLCB22.49 [imported] - *Mycobacterium leprae* (fragment)
C;Species: *Mycobacterium leprae*
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T44914
R;Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z22864
A;Accession: T44914
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-107 <PAR>
A;Cross-references: UNIPROT:Q32980; EMBL:Z98741; PIDN:CAB11414.1
A;Experimental source: cosmid B22
C;Genetics:
A;Note: MLCB22.49

Query Match 63.2%; Score 36; DB 2; Length 107;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 11
Db 54 TSSMPEVRRR 64
|:|:|

RESULT 11
T36913

probable integral membrane sugar transport protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C:Accession: T36913

R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, July 1999
 A:Reference number: Z21574

A:Accession: T36913
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-314 <SEE>

A:Cross-references: UNIPROT:Q9X9X3; EMBL:AL096743; PIDN:CAB46400.1; GSPDB:GN00070; SCOE

A:Experimental source: strain A3(2)

C:Genetics:

C:Superfamily: inner membrane protein ugpA

Query Match 63.2%; Score 36; DB 2; Length 314;
 Best Local Similarity 85.7%; Pred. No. 34;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TLPHRR 9

Db 225 TLPHRR 231

RESULT 12

S22151

gamma-1-microglobulin precursor - plaice

C:Species: Pleuronectes platessa (plaice)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: S22151

R:Leaver, M.J.
 submitted to the EMBL Data Library, December 1991

A:Reference number: S22151

A:Accession: S22151

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-355 <LEA>

A:Cross-references: UNIPROT:P36992; EMBL:X63762; NID:g64233; PIDN:CAA45294.1; PID:g64234

C:Superfamily: alpha-1-microglobulin and inter-alpha-trypsin inhibitor light chain precu

F:37-189/Domain: lipocalin homology <LIP>

F:236-286/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>

F:292-342/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>

Query Match 63.2%; Score 36; DB 1; Length 355;

Best Local Similarity 54.5%; Pred. No. 38;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TSTLPHRRTR 11

Db 52 TSSCPHQRNR 62

RESULT 13

C95053

IS1167, transposase [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004

C:Accession: C95053

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Hickey, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

son, T.; Umayam, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: C95053

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-422 <KUR>

A:Cross-references: UNIPROT:Q97SC5; GB:AB005672; PIDN:AAK74620.1; PID:gl4971930; GSPDB:G

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0460

C:Superfamily: Lactobacillus delbrueckii ISL3 transposase

Query Match 63.2%; Score 36; DB 2; Length 422;

Best Local Similarity 75.0%; Pred. No. 46;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TLPHRR 10

Db 342 TLPHRR 349

RESULT 14

T30645

hypothetical protein 43L - Molluscum contagiosum virus 1

N:Alternate names: MC043L

C:Species: Molluscum contagiosum virus 1

C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004

C:Accession: T30645

R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.

Science 273, 813-816, 1996

A>Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host res

A:Reference number: Z20876; MUID:96325459; PMID:8670425

A:Accession: T30645

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-772 <SEN>

A:Cross-references: UNIPROT:Q98211; EMBL:U60315; NID:gl491943; PIDN:AAC55171.1; PID:gl491

C:Genetics:

A>Note: MC043L

C:Superfamily: Molluscum contagiosum virus 1 hypothetical protein 43L

Query Match 63.2%; Score 36; DB 2; Length 772;

Best Local Similarity 70.0%; Pred. No. 86;

Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 STLPHRRTR 11

Db 470 SPLPHARRVR 479

RESULT 15

JQ2205

UL47h protein - Marek's disease virus

C:Species: Marek's disease virus

C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004

C:Accession: JQ2205

R:fanagida, N.; Yoshida, S.; Nazerian, K.; Lee, L.F.

J. Gen. Virol. 74, 1837-1845, 1993

A>Title: Nucleotide and predicted amino acid sequences of Marek's disease virus homologues

A:Reference number: JQ2199; MUID:93389438; PMID:8397281

A:Accession: JQ2205

A:Molecule type: DNA

A:Residues: 1-808 <YAN>

A:Cross-references: UNIPROT:Q69320; GB:L10283; NID:G388703; PIDN:AAA03150.1; PID:G388708

A:Experimental source: strain GA

Query Match 63.2%; Score 36; DB 2; Length 808;

Best Local Similarity 62.5%; Pred. No. 90;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 STLPHRR 9

Db 451 STMPHVRX 458

Search completed: November 10, 2004, 13:40:42

Job time: 3.45078 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 13.4318 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-34

Perfect score: 57

Sequence: 1 TSLPHIRTR 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	43	75.4	615	2 Q93L36	Q93L36 rhizobium 1
2	42	73.7	465	2 Q6CG25	Q6CG25 yarrowia 11
3	42	73.7	1114	2 Q6RYH8	Q6RYH8 cryptococcus
4	42	73.7	1114	2 AAS67509	AAS67509 cryptococcus
5	41	71.9	402	2 Q6Z4N7	Q6Z4N7 oryza sativ
6	41	71.9	402	2 BAC83803	BAC83803 oryza sat
7	40	70.2	339	2 Q8Q2E1	Q8Q2E1 streptomyces
8	40	70.2	870	2 Q22822	Q22822 caenorhabdit
9	39	68.4	762	2 Q22587	Q22587 nicotiana t
10	38.5	67.5	201	2 Q9XJ17	Q9XJ17 oryza sativ
11	38	66.7	42	2 Q9PDM8	Q9PDM8 xylella fas
12	38	66.7	129	2 Q6JEH8	Q6JEH8 debaryomyces
13	38	66.7	129	2 AAT12528	AAT12528 debaryomy
14	38	66.7	199	2 Q7Q5V1	Q7Q5V1 anopheles g
15	38	66.7	301	2 Q7L7C4	Q7L7C4 nectria cin
16	38	66.7	301	2 AAQ11957	AAQ11957 nectria c
17	38	66.7	308	2 Q6FEB7	Q6FEB7 acinetobact
18	38	66.7	320	2 Q6JEG5	Q6JEG5 candida dub
19	38	66.7	320	2 AAT12541	AAT12541 candida d
20	38	66.7	330	2 Q6JEI0	Q6JEI0 pichia jadi
21	38	66.7	330	2 AAT12526	AAT12526 pichia ja
22	38	66.7	347	2 Q6JEG3	Q6JEG3 pichia gull
23	38	66.7	347	2 AAT12543	AAT12543 pichia gu
24	38	66.7	348	2 Q6JEH6	Q6JEH6 candida nor
25	38	66.7	348	2 AAT12530	AAT12530 candida n
26	38	66.7	350	2 Q6JEH5	Q6JEH5 candida ces
27	38	66.7	350	2 AAT12531	AAT12531 candida c
28	38	66.7	351	2 Q6DN96	Q6DN96 candida mal
29	38	66.7	352	2 Q6JEG4	Q6JEG4 candida tro
30	38	66.7	352	2 AAT12542	AAT12542 candida t
31	38	66.7	353	2 Q6H191	Q6H191 debaryomyces

ALIGNMENTS

RESULT 1

Q93L36 PRELIMINARY; PRT; 615 AA.
 ID Q93L36
 AC Q93L36;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Rhizobium leguminosarum (biovar viciae).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 OX NCBI_TaxID=387;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF39;
 RX MEDLINE=99113394; PubMed=9914965;
 RA Sadykov M.R., Ivashina I.V., Kanapin A.A., Shliapnikov M.G.,
 RA Ksenzenko V.N.;
 RA "Structure-functional organization of exopolysaccharide biosynthetic genes in Rhizobium leguminosarum bv. viciae VF39.";
 RL Mol. Biol. (Mosk.) 32:797-804(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF39;
 RA Ivashina T.V., Sadykov M.R., Kalinchuk N.A., Senchenkova S.N.,
 RA Druzhinina T.N., Grachev A.A., Shashkov A.S., Shibaev V.N.,
 RA Kanapin A.A., Shlyapnikov M.G., van Workum W.A., Kijne J.W.,
 RA Ksenzenko V.N.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF028810; AAK77321.1;
 DR InterPro; IPR001440; TPR.
 DR InterPro; IPR008941; TPR-like.
 DR PROSITE; PS50293; TPR_REGION; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 615 AA; 68091 MW; EF24ABAB32C1DB04E CRC64;

Query Match 75.4%; Score 43; DB 2; Length 615;

Best Local Similarity 88.9%; Pred.No. 15;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TLPHIRTR 11

DB 389 TLPHIRRR 397

RESULT 2

Q6CG25 PRELIMINARY; PRT; 465 AA.
 ID Q6CG25
 AC Q6CG25;
 DT 01-OCT-2004 (TrEMBLrel. 28, Created)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Similar to CA2749|IPF11465 Candida albicans IPF11465 unknown function.

GN ORFNames-YALIOB01452g;
 OS Yarrowia lipolytica (Candida lipolytica).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Dipodascaceae; Yarrowia.
 OX NCBI_TaxID=4952;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIB99;
 RG GENOEUVRES;
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
 RA Goffard N., Frangeul L., Aigle M., Anichouard V., Babour A., Barbe V.,
 RA Barney S., Blanchin S., Seckerich J.M., Bayne E., Bleykasten C.,
 RA Boisdrame A., Boyer J., Catolico L., Confanioleri F., de Barvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.M., Nikolski M., Oztas S., Ozler-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
 RA Swennene D., Tekai F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RA Winkler P., Souciet J.L.;
 RT "Genome evolution in yeasts."
 RL Nature 430:35-44(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIB99;
 RA Genoscope;
 RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 RL EMBL; CR382128; CAG82604.1; -
 SQ SEQUENCE 465 AA; 52869 MW; 1D56E4B4CBBC88 CRC64;

 Query Match 73.7%; Score 42; DB 2; Length 465;
 Best Local Similarity 63.6%; Pred. No. 17;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

 QY 1 TSTLPPIRRTR 11
 |||:|||||
 Db 174 TSSIPVHVRFR 184

 RESULT 3
 ID Q6RYH8 PRELIMINARY; PRT; 1114 AA.
 AC Q6RYH8;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE DNA-dependent RNA polymerase II second largest subunit
 DE (Fragment).
 OS Cryptococcus neoformans (Filobasidiella neoformans).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
 OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
 OX NCBI_TaxID=5207;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC PubMed=15070748;
 RA Liu Y.J., Hall B.D.;
 RT "Body plan evolution of ascomycetes, as inferred from an RNA
 polymerase II phylogeny."
 RL Proc. Natl. Acad. Sci. U.S.A. 101:4507-4512(2004).
 DR EMBL; AY485620; AAS67509.1; -
 FT NON TER 1114 1114
 SQ SEQUENCE 1114 AA; 125436 MW; C74396A8DF649CBF CRC64;

 Query Match 73.7%; Score 42; DB 2; Length 1114;
 Best Local Similarity 80.0%; Pred. No. 46;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 1 TSTLPPIRRTR 10
 |||:|||||
 Db 505 TSTLSHLRRT 514

 RESULT 5
 ID Q6Z4N7 PRELIMINARY; PRT; 402 AA.
 AC Q6Z4N7;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative r40c1 protein.
 GN Name=OSJNBa0060017.9;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sasaki T., Matsumoto T., Katayose Y.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005167; BAC38303.1;
 DR InterPro; IPR008997; RcinB_like.

DR Pfam; PF04563; RNA_pol_Rpb2_1; 1.
 DR Pfam; PF04561; RNA_pol_Rpb2_2; 1.
 DR Pfam; PF04565; RNA_pol_Rpb2_3; 1.
 DR Pfam; PF04566; RNA_pol_Rpb2_4; 1.
 DR Pfam; PF04567; RNA_pol_Rpb2_5; 1.
 DR Pfam; PF00562; RNA_pol_Rpb2_6; 1.
 DR PROSITE; PS01166; RNA_POL_BETA; 1.
 KW DNA-directed RNA polymerase; Transcription; Transferase.
 FT NON TER 1114 1114
 SQ SEQUENCE 1114 AA; 125436 MW; C74396A8DF649CBF CRC64;

 Query Match 73.7%; Score 42; DB 2; Length 1114;
 Best Local Similarity 80.0%; Pred. No. 46;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 1 TSTLPPIRRTR 10
 |||:|||||
 Db 505 TSTLSHLRRT 514

 RESULT 4
 ID AAS67509 PRELIMINARY; PRT; 1114 AA.
 AC AAS67509;
 DT 14-APR-2004 (TrEMBLrel. 27, Created)
 DT 14-APR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 14-APR-2004 (TrEMBLrel. 27, Last annotation update)
 DE DNA-dependent RNA polymerase II second largest subunit
 DE (Fragment).
 OS Cryptococcus neoformans (Filobasidiella neoformans).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
 OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
 OX NCBI_TaxID=5207;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC PubMed=15070748; B.D.;
 RA Liu Y.J., Hall B.D.;
 RT "Body plan evolution of ascomycetes, as inferred from an RNA
 polymerase II phylogeny."
 RL Proc. Natl. Acad. Sci. U.S.A. 101:4507-4512(2004).
 DR EMBL; AY485620; AAS67509.1; -
 FT NON TER 1114 1114
 SQ SEQUENCE 1114 AA; 125436 MW; C74396A8DF649CBF CRC64;

 Query Match 73.7%; Score 42; DB 2; Length 1114;
 Best Local Similarity 80.0%; Pred. No. 46;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 1 TSTLPPIRRTR 10
 |||:|||||
 Db 505 TSTLSHLRRT 514

 RESULT 5
 ID Q6Z4N7 PRELIMINARY; PRT; 402 AA.
 AC Q6Z4N7;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative r40c1 protein.
 GN Name=OSJNBa0060017.9;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sasaki T., Matsumoto T., Katayose Y.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005167; BAC38303.1;
 DR InterPro; IPR008997; RcinB_like.

SQ SEQUENCE 402 AA; 44455 MW; 667C4B59F72E7589 CRC64;
 Query Match 71.9%; Score 41; DB 2; Length 402;
 Best Local Similarity 63.6%; Pred. No. 23;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPHRRTR 11
 :|:|:|:|:
 Db 200 TATTPHLRSTR 210

RESULT 6
 BAC383803 PRELIMINARY; PRT; 402 AA.
 AC BAC383803;
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative r40c1 protein.
 GN OSJNBA0060017.9.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza; Oryza sativa.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Katayose Y.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
 clone:OSJNBA0060017.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005167; BAC383803.1; -
 SQ SEQUENCE 402 AA; 44455 MW; 667C4B59F72E7589 CRC64;

Query Match 71.9%; Score 41; DB 2; Length 402;
 Best Local Similarity 63.6%; Pred. No. 23;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPHRRTR 11
 :|:|:|:|:
 Db 200 TATTPHLRSTR 210

RESULT 7
 BAC383803 PRELIMINARY; PRT; 339 AA.
 AC BAC383803;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocNames=SAV568;
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4580;
 RX MEDLINE=2147403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 avermitilis: deducing the ability of producing secondary
 metabolites.";
 RL Proc. Nati. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4580;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,

RA Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AF005023; BAC68278.1; -
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR001932; PP2C-like.
 DR Pfam; PF02518; HATPase_c; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR SMART; SM00331; PP2C_SIG; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 339 AA; 86011 MW; 0D3A5258234F898F CRC64;

Query Match 70.2%; Score 40; DB 2; Length 339;
 Best Local Similarity 70.0%; Pred. No. 30;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STLPHRRTR 11
 :|:|:|:|:
 Db 275 STTPLRRAR 284

RESULT 8
 Q22822 PRELIMINARY; PRT; 870 AA.
 AC Q22822;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein T27A10.6.
 GN CrfNames=T27A10.6;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Wilson R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Stellyes L.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Wilson R.;
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U58734; AB52504.3; -
 DR PIR; F89500; F89500.
 DR WormPep; T27A10.6; CE30193.
 KW Hypothetical protein.
 SQ SEQUENCE 870 AA; 98207 MW; B3AC8105E0308A45 CRC64;

Query Match 70.2%; Score 40; DB 2; Length 870;
 Best Local Similarity 72.7%; Pred. No. 85;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSTLPHRRTR 11
 :|:|:|:|:
 Db 524 TKTTPRRTR 534

RESULT 9
ID O22587 PRELIMINARY; PRT; 762 AA.
AC O22587;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ethylene receptor homolog.
GN Name=ETH1;
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Nicotiana.
OC NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang J.-S., Xie C., Liu P., Liu F., Chen S.-Y.;
RT "A novel tobacco gene coding for a product similar to bacterial two-
RT component regulators."
RL Chin. Sci. Bull. 44:1025-1029(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Zhang J.S., Xie C., Shen Y.G., Chen S.Y.;
RT "A two-component gene (ETH1) encoding a putative ethylene receptor
RT homolog is both developmentally and stress-regulated in tobacco."
RL Theor. Appl. Genet. 102:815-824(2001).
RN [3]
RP SEQUENCE FROM N.A.
EX MEDLINE=22149310; PubMed=12154144;
RA Xie C., Zhang Z.G., Zhang J.S., He X.J., Cao W.H., He S.J., Chen S.Y.;
RT "Spatial expression and characterization of a putative ethylene
RT receptor protein ETH1 in tobacco."
RL Plant Cell Physiol. 43:810-815(2002).
RN [4]
RP SEQUENCE FROM N.A.
EX MEDLINE=22424050; PubMed=12535351;
RA Xie C., Zhang J.S., Zhou H.L., Li J., Zhang Z.G., Wang D.W.,
RA Chen S.Y.;
RT "Serine/threonine kinase activity in the putative histidine kinase-
RT like ethylene receptor ETH1 from tobacco."
RL Plant J. 33:385-393(2003).
CC -1- SIMILARITY: Contains 1 histidine kinase domain.
DR EVBL; AF026267; AAC31213.3; -
DR PIR; T01930; T01930.
DR HSP; P49333; IBCF.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:000527; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0000156; F:two-component response regulator activity; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0000150; P:two-component signal transduction system (p...); IEA.
DR InterPro; IPR003594; APBbind ATPase.
DR InterPro; IPR011006; Chey_like.
DR InterPro; IPR003018; GAF.
DR InterPro; IPR005467; His_kinase.
DR InterPro; IPR003661; His_kin_N.
DR InterPro; IPR009082; His_kin_Nomdm.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF01590; GAF; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00512; Hisk1; 1.
DR Pfam; PF00072; Response_reg; 1.
DR ProDom; PD000039; Response_reg; 1.
DR SMART; SM00065; GAF; 1.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00388; Hisk1; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS50109; HIS_KIN; 1.
DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.

KW Phosphorylation; Receptor; Sensory transduction.
SQ SEQUENCE 762 AA; 85320 MW; F9071592179EA3CE CRC64;

Query Match 68.4%; Score 39; DB 2; Length 762;
Best Local Similarity 72.7%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSTLPPIRRT 11
|||
Db 554 TSTLPCCCTR 564

RESULT 10

Q9XJ17
ID Q9XJ17 PRELIMINARY; PRT; 201 AA.
AC Q9XJ17;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OC NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Nagamura Y., Yamamoto K.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB023482; BAA78735.1; -
DR Gramene; Q9XJ17; -
KW Hypothetical protein.
SQ SEQUENCE 201 AA; 21230 MW; DDE198727295403 CRC64;

Query Match 67.5%; Score 38.5; DB 2; Length 201;
Best Local Similarity 56.2%; Pred. No. 32;
Matches 9; Conservative 2; Mismatches 0; Indels 5; Gaps 1;

QY 1 TSTLPPI----RRT 11
|||
Db 97 TATVPHIGRAVPRTR 112

RESULT 11

Q9PDM8
ID Q9PDM8 PRELIMINARY; PRT; 42 AA.
AC Q9PDM8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=Xf1351;
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OC NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfy H.,
RA Pacinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Prohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,

```

RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Nhani A.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.B., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Teai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Stubal J.C.,
RT "the genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE003967; AAF84160.1; -.
DR PIR; F82691; F82691.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 42 AA; 4915 MW; 0F9E51F876FE0C9E CRC64;

Query Match 66.7%; Score 38; DB 2; Length 42;
Best Local Similarity 70.0%; Pred. No. 7;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STLPHIRRT 11
Db 17 STWPHQHRT 26

RESULT 12
Q6JEH8 PRELIMINARY; PRT; 129 AA.
AC Q6JEH8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 08-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE RNA polymerase II second largest subunit (fragment).
OS Debaryomyces carsonii.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=28549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 2285;
RA Diezmann S., Cox C.J., Schoenian G., Vilgalys R.J., Mitchell T.G.;
RL Submitted (DDB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY497608; AAT12528.1; -.
DR InterPro; IPR007645; RNA_pol_Rpb2_3.
DR Pfam; PF04565; RNA_pol_Rpb2_3; 1.
FT NON_TER 129
SQ SEQUENCE 129 AA; 14706 MW; 82AEC0B1D7D28326 CRC64;

Query Match 66.7%; Score 38; DB 2; Length 129;
Best Local Similarity 70.0%; Pred. No. 25;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
Db 86 SSTLSHLRRT 95

RESULT 13
AAT12528 PRELIMINARY; PRT; 129 AA.
AC AAT12528;
DT 20-MAY-2004 (TrEMBLrel. 27, Created)
DT 20-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DT 20-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE RNA polymerase II second largest subunit (fragment).
OS Debaryomyces carsonii.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.

RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Nhani A.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.B., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Teai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Stubal J.C.,
RT "the genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE003967; AAF84160.1; -.
DR PIR; F82691; F82691.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 42 AA; 4915 MW; 0F9E51F876FE0C9E CRC64;

Query Match 66.7%; Score 38; DB 2; Length 129;
Best Local Similarity 70.0%; Pred. No. 25;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
Db 86 SSTLSHLRRT 95

RESULT 14
Q7Q5V1 PRELIMINARY; PRT; 199 AA.
AC Q7Q5V1;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AGCP5849 (Fragment).
GN Name=agCG52391; ORFNames=ENSMANG00000018149;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008960; EAA11578.1; -.
FT NON_TER 1
SQ SEQUENCE 199 AA; 22656 MW; 31C0D715B1BC11C0 CRC64;

Query Match 66.7%; Score 38; DB 2; Length 199;
Best Local Similarity 70.0%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSTLPHIRRT 10
Db 59 TSNLQHVRT 68

RESULT 15
Q717C4 PRELIMINARY; PRT; 301 AA.
AC Q717C4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE RNA polymerase II subunit (fragment).
GN Name=RPB2;
OS Nectria cinnabarina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Nectria.
OX NCBI_TaxID=29855;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GJS 91-111;

```

RA Chaverri P., Castlebury L.A., Overton B.E., Samuels G.J.;
RT "Hypocrea/Trichoderma: species with conidiophore elongations and green
RL conidia."; 95:1100-1140(2003).

KL Mycologia 95:1100-1140(2003).
DR EMBL; AF545567; AAQ11957.1; -.
DR InterPro; IPR007645; RNA_pol_Rpb2_3.
DR InterPro; IPR007646; RNA_pol_Rpb2_4.
DR InterPro; IPR007647; RNA_pol_Rpb2_5.
DR Pfam; PF04565; RNA_pol_Rpb2_3; 1.
DR Pfam; PF04566; RNA_pol_Rpb2_4; 1.
DR Pfam; PF04567; RNA_pol_Rpb2_5; 1.
FT NON TER 1
FT NON TER 301
SQ SEQUENCE 301 AA; 33976 MW; C1942A7C034E2B1C CRC64;

Query Match 66.7%; Score 38; DB 2; Length 301;
Best Local Similarity 70.0%; Pred.No. 64;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TSTLPHIRRT 10
Db 24 SSTLSHLRRT 33

Search completed: November 10, 2004, 13:38:34
Job time : 14.4318 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein, - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 4.14508 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-35

Perfect score: 60

Sequence: 1 NGNLFASFADS 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	70.0	226	3	US-09-097-889-18
2	42	70.0	226	4	US-09-098-079-18
3	42	70.0	404	4	US-09-248-796A-19544
4	40	66.7	207	4	US-08-811-519-25
5	40	66.7	485	1	US-08-453-956-15
6	40	66.7	485	1	US-08-086-631-15
7	40	66.7	485	2	US-08-452-930-15
8	40	66.7	485	5	PCT-US93-08174-15
9	39	65.0	116	4	US-09-489-039A-13377
10	39	65.0	120	4	US-09-270-767-61086
11	39	65.0	148	4	US-09-270-767-45573
12	37.5	62.5	311	4	US-09-248-796A-17937
13	36	60.0	377	4	US-09-328-352-4688
14	36	60.0	595	4	US-09-198-452A-74
15	35	58.3	330	4	US-09-540-236-2570
16	35	58.3	712	1	US-08-121-713D-64
17	35	58.3	712	1	US-08-835-268-64
18	35	58.3	712	2	US-09-060-692-64
19	35	58.3	712	3	US-08-833-391-64
20	35	58.3	712	3	US-09-060-610-64
21	35	58.3	712	5	PCT-US94-10151A-64
22	34	56.7	162	1	US-08-453-956-18
23	34	56.7	162	1	US-08-086-631-18
24	34	56.7	162	2	US-08-452-930-18
25	34	56.7	162	5	PCT-US93-08174-18
26	34	56.7	222	3	US-08-845-346-10
27	34	56.7	294	1	US-08-142-439A-4

28	34	56.7	294	2	US-08-869-477-4	Sequence 4, Appli
29	34	56.7	445	4	US-09-328-352-7248	Sequence 7248, Ap
30	34	56.7	463	1	US-08-142-439A-2	Sequence 2, Appli
31	34	56.7	463	2	US-08-869-477-2	Sequence 2, Appli
32	34	56.7	463	4	US-09-631-603-16	Sequence 16, Appli
33	34	56.7	477	1	US-08-453-956-25	Sequence 25, Appli
34	34	56.7	477	1	US-08-086-631-25	Sequence 25, Appli
35	34	56.7	477	2	US-08-452-930-25	Sequence 17, Appli
36	34	56.7	477	4	US-09-631-603-17	Sequence 17, Appli
37	34	56.7	477	5	PCT-US93-08174-25	Sequence 25, Appli
38	34	56.7	509	3	US-08-845-346-2	Sequence 2, Appli
39	34	56.7	553	3	US-08-845-346-12	Sequence 12, Appli
40	33	55.0	166	4	US-09-583-110-4577	Sequence 4577, Ap
41	33	55.0	541	4	US-09-328-352-7237	Sequence 7237, Ap
42	33	55.0	556	3	US-08-687-590-24	Sequence 24, Appli
43	33	55.0	556	4	US-09-328-352-7560	Sequence 7560, Ap
44	33	55.0	750	6	5457037-3	Patent No. 5457037
45	33	55.0	751	6	5457037-5	Patent No. 5457037

ALIGNMENTS

RESULT 1

US-09-097-889-18

; Sequence 18, Application US/09097889

; Patent No. 6218117

; GENERAL INFORMATION:

; APPLICANT: Hertnstadt, Corrina

; APPLICANT: Ghosh, Soumitra S.

; APPLICANT: Davis, Robert E.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING

; TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE

; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/097,889

; FILING DATE: 15-JUN-1998

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Rosenman Ph.D., Stephen J.

; REGISTRATION NUMBER: 43,058

; REFERENCE/DOCKET NUMBER: 660088.417

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 226 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; US-09-097-889-18

Query Match 70.0%; Score 42; DB 3; Length 226;

Best Local Similarity 90.0%; Pred. No. 6.3;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFAD 10

DB 2 NGNLFASFAD 11

```

; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19544

Query Match      70.0%; Score 42; DB 4; Length 226;
Best Local Similarity 90.0%; Pred. No. 6.3;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 NGNLFASFIA 10
        |||||
DB      2 NGNLFASFIA 11

RESULT 3
US-09-248-796A-19544
; Sequence 19544, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19544
; LENGTH: 404

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,079
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman Ph.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 660088.416
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-098-079-18

Query Match      70.0%; Score 42; DB 4; Length 226;
Best Local Similarity 90.0%; Pred. No. 6.3;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 NGNLFASFIA 10
        |||||
DB      2 NGNLFASFIA 11

RESULT 3
US-09-248-796A-19544
; Sequence 19544, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19544
; LENGTH: 404

```

```

; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19544

Query Match      70.0%; Score 42; DB 4; Length 404;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 NGNLFASPIADS 12
        |||||
DB      2.9 NGNLFASVSEDS 230

RESULT 4
US-08-811-519-25
; Sequence 25, Application US/08811519B
; Patent No. 6630345
; GENERAL INFORMATION:
; APPLICANT: Petrenko, Alexandre
; TITLE OF INVENTION: CALCIUM INDEPENDENT RECEPTOR OF ALPHA-LATROTOXIN,
; FILE REFERENCE: 1049-1-007
; CURRENT APPLICATION NUMBER: US/08/811,519B
; CURRENT FILING DATE: 1997-03-04
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 207
; TYPE: PRT
; ORGANISM: rat
US-08-811-519-25

Query Match      66.7%; Score 40; DB 4; Length 207;
Best Local Similarity 77.8%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 NGNLFASPI 9
        |||||
DB      36 HGNLFASFV 44

RESULT 5
US-08-453-956-15
; Sequence 15, Application US/08453956
; Patent No. 5770445
; GENERAL INFORMATION:
; APPLICANT: Kindvogel, Wayne R.
; TITLE OF INVENTION: GLUCAGON RECEPTORS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 COLUMBIA CENTER
; CITY: SEATTLE
; STATE: WA
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,956
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,631
; FILING DATE: July 1, 1993
; APPLICATION NUMBER: US 07/938,331
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963

```

US-08-452-930-15
; Sequence 15, Application US/08452930
; Patent No. 5919635
; GENERAL INFORMATION:
; APPLICANT: Kindvogel, Wayne R.
; TITLE OF INVENTION: GLUCAGON RECEPTORS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 COLUMBIA CENTER
; CITY: SEATTLE
; STATE: WA
; COUNTRY: USA
; ZIP: 99104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,930
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,631
; FILING DATE: July 1, 1993
; APPLICATION NUMBER: US 07/938,331
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 990008.424C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-452-930-15
Query Match 66.7%; Score 40; DB 1; Length 485;
Best Local Similarity 77.8%; Pred. No. 33;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 NGNLFASFI 9
Db 178 HGNLFASV 186
RESULT 6
US-08-086-631-15
; Sequence 15, Application US/08086631
; Patent No. 5776725
; GENERAL INFORMATION:
; APPLICANT: Kindvogel, Wayne R.
; TITLE OF INVENTION: GLUCAGON RECEPTORS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 COLUMBIA CENTER
; CITY: SEATTLE
; STATE: WA
; COUNTRY: USA
; ZIP: 99104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,631
; FILING DATE: July 1, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938,331
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 990008.424C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-086-631-15
Query Match 66.7%; Score 40; DB 1; Length 485;
Best Local Similarity 77.8%; Pred. No. 33;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 NGNLFASFI 9
Db 178 HGNLFASV 186
RESULT 7

US-08-452-930-15
; Sequence 15, Application US/08452930
; Patent No. 5919635
; GENERAL INFORMATION:
; APPLICANT: Kindvogel, Wayne R.
; TITLE OF INVENTION: GLUCAGON RECEPTORS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 COLUMBIA CENTER
; CITY: SEATTLE
; STATE: WA
; COUNTRY: USA
; ZIP: 99104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,930
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,631
; FILING DATE: July 1, 1993
; APPLICATION NUMBER: US 07/938,331
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 990008.424C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-452-930-15
Query Match 66.7%; Score 40; DB 2; Length 485;
Best Local Similarity 77.8%; Pred. No. 33;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 NGNLFASFI 9
Db 178 HGNLFASV 186
RESULT 8
PCT-US93-08174-15
; Sequence 15, Application PC/TUS9308174
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: NAME: ZymoGenetics, Inc.
; APPLICANT: STREET: 4225 Roosevelt Way North East
; APPLICANT: CITY: Seattle, Washington
; APPLICANT: COUNTRY: United States
; APPLICANT: POSTAL CODE: 98105
; APPLICANT: TELEPHONE: (206) 547-80808
; TITLE OF INVENTION: GLUCAGON RECEPTORS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 COLUMBIA CENTER
; CITY: SEATTLE
; STATE: WA
; COUNTRY: USA
; ZIP: 99104-7092
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/08174
 ; FILING DATE: 30-AUG-1993
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/086,631
 ; FILING DATE: July 1, 1993
 ; APPLICATION NUMBER: US 07/938,331
 ; FILING DATE: 28-AUG-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McMasters, David D.
 ; REGISTRATION NUMBER: 33,963
 ; REFERENCE/DOCKET NUMBER: 990008.424C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-622-4900
 ; TELEFAX: 206-682-6031
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 485 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-MUS93-08174-15

Query Match 66.7%; Score 40; DB 5; Length 485;
 Best Local Similarity 77.8%; Pred. No. 33;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NGNLFASFI 9
 DB 178 HGNLFASFV 186

RESULT 9
 US-09-489-039A-13377
 ; Sequence 13377, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 13377
 ; LENGTH: 116
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-13377

Query Match 65.0%; Score 39; DB 4; Length 116;
 Best Local Similarity 58.3%; Pred. No. 10;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NGNLFASFIADS 12
 DB 49 SGTAFASFVGD 60

RESULT 10
 US-09-270-767-61086
 ; Sequence 61086, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 61086
 ; LENGTH: 120
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; US-09-270-767-61086

Query Match 65.0%; Score 39; DB 4; Length 120;
 Best Local Similarity 80.0%; Pred. No. 11;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NLFAFIADS 12
 DB 13 NLFAFIADS 22

RESULT 11
 US-09-270-767-45573
 ; Sequence 45573, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 45573
 ; LENGTH: 148
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; FEATURE:
 ; OTHER INFORMATION: Xaa means any amino acid
 ; US-09-270-767-45573

Query Match 65.0%; Score 39; DB 4; Length 148;
 Best Local Similarity 80.0%; Pred. No. 13;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NLFAFIADS 12
 DB 41 NLFAFIADS 50

RESULT 12
 US-09-248-796A-17937
 ; Sequence 17937, Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstock et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
 ; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; CURRENT FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; PRIOR FILING DATE: 1998-08-13
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 17937
 ; LENGTH: 311
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 ; US-09-248-796A-17937

Query Match 62.5%; Score 37.5; DB 4; Length 311;
 Best Local Similarity 83.3%; Pred. No. 56;

Matches 10; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 NGNLFASFIADS 12
| | | | | | | | | |
Db 66 NANLFAS-IADS 76

RESULT 13

US-09-328-352-4688
; Sequence 4688, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4688
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4688

Query Match 60.0%; Score 36; DB 4; Length 377;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFI 9
: | | | | | | | | | |
Db 199 HGNIKXSF 207

RESULT 14

US-09-198-452A-74
; Sequence 74, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 74
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-74

Query Match 60.0%; Score 36; DB 4; Length 595;
Best Local Similarity 63.6%; Pred. No. 2.1e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 NGNLFASFIAD 11
| | | | | | | | | |
Db 345 NGNQMASILAD 355

RESULT 15

US-09-540-236-2570
; Sequence 2570, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2570
; LENGTH: 330
; TYPE: PRT
; ORGANISM: M.catarhalis
US-09-540-236-2570

Query Match 58.3%; Score 35; DB 4; Length 330;
Best Local Similarity 60.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GNLFASFIAD 11
| | | | | | | | | |
Db 117 GNSFSSFVRD 126

Search completed: November 10, 2004, 13:44:07
Job time : 5.14508 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 12.8912 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-35

Perfect score: 60

Sequence: 1 NGNLFASFIADS 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	100.0	12	14	US-10-092-750-35
2	42	70.0	53	14	Sequence 35, Appl
3	42	70.0	58	14	Sequence 5361, Ap
4	42	70.0	226	9	Sequence 6592, Ap
5	42	70.0	226	16	Sequence 18, Appl
6	40	66.7	485	13	Sequence 192, App
7	38	63.3	79	17	Sequence 2, Appl
8	38	63.3	85	17	Sequence 186935,
9	38	63.3	103	16	Sequence 362727,
10	38	63.3	222	15	Sequence 141470,
11	38	63.3	294	15	Sequence 52549, A
12	38	63.3	327	15	Sequence 2, Appl
13	38	63.3	527	15	Sequence 52852, A
					Sequence 43548, A

14	38	63.3	527	17	US-10-739-930-8886	Sequence 8886, Ap
15	38	63.3	645	15	US-10-424-599-210244	Sequence 210244,
16	38	63.3	691	15	US-10-424-599-209563	Sequence 209563,
17	38	63.3	1019	16	US-10-408-765A-2344	Sequence 2344, Ap
18	38	63.3	1874	16	US-10-437-963-135735	Sequence 135735,
19	38	63.3	1887	15	US-10-282-122A-48562	Sequence 48562, A
20	37.5	62.5	564	14	US-10-032-585-7313	Sequence 7313, Ap
21	37	61.7	339	15	US-10-451-337-10	Sequence 10, Appl
22	37	61.7	402	15	US-10-282-122A-70412	Sequence 70412, A
23	37	61.7	540	15	US-10-424-599-169281	Sequence 169281,
24	37	61.7	707	16	US-10-437-963-160125	Sequence 160125,
25	37	61.7	707	16	US-10-437-963-199032	Sequence 199032,
26	36	60.0	42	17	US-10-425-115-367795	Sequence 367795,
27	36	60.0	54	16	US-10-437-963-163310	Sequence 163310,
28	36	60.0	61	9	US-09-925-388-751	Sequence 751, App
29	36	60.0	61	14	US-10-102-806-751	Sequence 751, App
30	36	60.0	94	9	US-09-925-398-757	Sequence 757, App
31	36	60.0	94	14	US-10-102-806-757	Sequence 757, App
32	36	60.0	128	17	US-10-425-115-351112	Sequence 351112,
33	36	60.0	149	16	US-10-767-701-60716	Sequence 60716, A
34	36	60.0	165	17	US-10-425-115-208835	Sequence 208835,
35	36	60.0	168	16	US-10-767-701-41984	Sequence 41984, A
36	36	60.0	204	15	US-10-424-599-214913	Sequence 214913,
37	36	60.0	257	15	US-10-371-592-1	Sequence 1, Appl
38	36	60.0	310	16	US-10-767-701-43500	Sequence 43500, A
39	36	60.0	399	17	US-10-425-115-291781	Sequence 291781,
40	36	60.0	403	17	US-10-425-115-291776	Sequence 291776,
41	36	60.0	419	15	US-10-425-114-47512	Sequence 47512, A
42	36	60.0	426	15	US-10-425-114-41019	Sequence 41019, A
43	36	60.0	455	16	US-10-437-963-202044	Sequence 202044,
44	36	60.0	499	15	US-10-425-114-65399	Sequence 65399, A
45	36	60.0	584	9	US-09-902-525-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1

US-10-092-750-35
; Sequence 35, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-35

Query Match 100.0%; Score 60; DB 14; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NGNLFASFIADS 12
DB 1 NGNLFASFIADS 12

RESULT 2

US-10-106-698-5361
; Sequence 5361, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:

```

; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA003P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5361
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6592

Query Match          70.0%; Score 42; DB 14; Length 53;
Best Local Similarity 90.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NGNLFASFIA 10
Db 8 NENLFASFIA 17

RESULT 3
US-10-106-698-6592
; Sequence 6592, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA003P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6592
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6592

Query Match          70.0%; Score 42; DB 14; Length 58;
Best Local Similarity 90.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NGNLFASFIA 10
Db 13 NENLFASFIA 22

RESULT 4
US-09-098-079-18
; Sequence 18, Application US/09098079
; Patent No. US20020064773A1
; GENERAL INFORMATION:
; APPLICANT: Herristadt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Cleveland, William
; APPLICANT: Fahy, Eoin F.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON QUANTIFICATION OF
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA

```

```

; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,079
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman Ph.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 660088.416
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-098-079-18

Query Match          70.0%; Score 42; DB 9; Length 226;
Best Local Similarity 90.0%; Pred. No. 10;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NGNLFASFIA 10
Db 2 NENLFASFIA 11

RESULT 5
US-10-408-765A-192
; Sequence 192, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-192

Query Match          70.0%; Score 42; DB 16; Length 226;
Best Local Similarity 90.0%; Pred. No. 10;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NGNLFASFIA 10
Db 2 NENLFASFIA 11

```

RESULT 6

US-10-010-065-2
; Sequence 2, Application US/10010065
; Publication No. US20020144300A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Matthews, William
; APPLICANT: Moore, Mark
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING GLUCAGON
; TITLE OF INVENTION: RECEPTOR GENE DISRUPTIONS
; FILE REFERENCE: R-648
; CURRENT APPLICATION NUMBER: US/10/010,065
; CURRENT FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: US 60/251,804
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: US 60/266,044
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2:
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-010-065-2

Query Match 56.7%; Score 40; DB 13; Length 485;
Best Local Similarity 77.8%; Pred. No. 57;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NGNLFASF1 9
:|||||:
DB 178 HGMLFASFV 186

RESULT 7

US-10-425-115-186935
; Sequence 186935, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 186935
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(79)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MMT4577_102073C.1.pep
US-10-425-115-186935

Query Match 63.3%; Score 38; DB 17; Length 79;
Best Local Similarity 88.9%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASF1 9
:|||||:
DB 2 NGNLFASF1 10

RESULT 8

US-10-425-115-362727
; Sequence 362727, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 362727
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(85)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MMT4577_93981C.1.pep
US-10-425-115-362727

Query Match 63.3%; Score 38; DB 17; Length 85;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NGNLFASF1ADS 12
:|||||:
DB 6 DGNLFGSFLSES 17

RESULT 9

US-10-437-963-141470
; Sequence 141470, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 141470
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_42570C.1.pep
US-10-437-963-141470

Query Match 63.3%; Score 38; DB 16; Length 103;
Best Local Similarity 77.8%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NGNLFASF1 9
:|||||:
DB 36 NGNLFVKFI 44

RESULT 10
US-10-425-114-52549

```

; Sequence 52549, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 52549
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3028-051-H9_FLI.pap
US-10-425-114-52549

Query Match      63.3%; Score 38; DB 15; Length 222;
Best Local Similarity 72.7%; Pred. No. 57;
Matches      8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 NGNLFASFIAD 11
DB      113 NGNLTSTFCAD 123

RESULT 11
US-10-371-592-2
; Sequence 2, Application US/10371592
; Publication No. US20040072774A1
; GENERAL INFORMATION:
; APPLICANT: COLUMBIA UNIVERSITY
; APPLICANT: Manfredi, Giovanni
; APPLICANT: Schon, Eric
; TITLE OF INVENTION: METHODS FOR EXPRESSING AND TARGETING MITOCHONDRIA-DNA-ENCODED
; FILE REFERENCE: 5199-6
; CURRENT APPLICATION NUMBER: US/10/371,592
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: US 60/358,935
; PRIOR FILING DATE: 2002-02-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-371-592-2

Query Match      63.3%; Score 38; DB 15; Length 294;
Best Local Similarity 100.0%; Pred. No. 78;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 NLFASFFIA 10
DB      67 NLFASFFIA 74

RESULT 12
US-10-425-114-52852
; Sequence 52852, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 52549
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3028-051-H9_FLI.pap
US-10-425-114-52549

Query Match      63.3%; Score 38; DB 15; Length 222;
Best Local Similarity 72.7%; Pred. No. 57;
Matches      8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 NGNLFASFIAD 11
DB      113 NGNLTSTFCAD 123

RESULT 11
US-10-371-592-2
; Sequence 2, Application US/10371592
; Publication No. US20040072774A1
; GENERAL INFORMATION:
; APPLICANT: COLUMBIA UNIVERSITY
; APPLICANT: Manfredi, Giovanni
; APPLICANT: Schon, Eric
; TITLE OF INVENTION: METHODS FOR EXPRESSING AND TARGETING MITOCHONDRIA-DNA-ENCODED
; FILE REFERENCE: 5199-6
; CURRENT APPLICATION NUMBER: US/10/371,592
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: US 60/358,935
; PRIOR FILING DATE: 2002-02-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-371-592-2

Query Match      63.3%; Score 38; DB 15; Length 294;
Best Local Similarity 100.0%; Pred. No. 78;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 NLFASFFIA 10
DB      67 NLFASFFIA 74

RESULT 12
US-10-425-114-52852
; Sequence 52852, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 52852
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3028-021-FLI_FLI.pap
US-10-425-114-52852

Query Match      63.3%; Score 38; DB 15; Length 327;
Best Local Similarity 72.7%; Pred. No. 88;
Matches      8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 NGNLFASFIAD 11
DB      218 NGNLTSTFCAD 228

RESULT 13
US-10-425-114-43548
; Sequence 43548, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43548
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700726816_FLI.pap
US-10-425-114-43548

Query Match      63.3%; Score 38; DB 15; Length 527;
Best Local Similarity 72.7%; Pred. No. 1.5e+02;
Matches      8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 NGNLFASFIAD 11
DB      418 NGNLTSTFCAD 428

RESULT 14
US-10-739-930-8886
; Sequence 8886, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 8886
; LENGTH: 527

```

```

; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: GLYMA-23APR03-C2174_9.p
US-10-739-930-8886

Query Match      63.3%; Score 38; DB 17; Length 527;
Best Local Similarity 72.7%; Pred. No. 1.5e+02;
Matches      8; Conservative      0; Mismatches      3; Indels      0; Gaps      0;

QY      1 NGNLFASFIAD 11
      |||||
Db      418 NGNLTSTFCAD 428

RESULT 15
US-10-424-599-210244
; Sequence 210244, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 210244
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(645)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_31879C.1.pep
US-10-424-599-210244

Query Match      63.3%; Score 38; DB 15; Length 645;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches      8; Conservative      0; Mismatches      3; Indels      0; Gaps      0;

QY      1 NGNLFASFIAD 11
      |||||
Db      536 NGNLTSTFCAD 546

```

Search completed: November 11, 2004, 01:28:17
Job time : 13.9412 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 2.67358 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-35
Perfect score: 60
Sequence: 1 NGNLFASFADS 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	42	70.0	226	1 PWH06	H+-transporting tw
2	42	70.0	226	2 F59153	H+-transporting tw
3	42	70.0	226	2 T10977	H+-transporting tw
4	40	66.7	226	2 T11498	H+-transporting tw
5	40	66.7	485	2 JC4363	glucagon receptor
6	40	66.7	485	2 JQ1957	glucagon receptor
7	38	63.3	226	1 FWH56	H+-transporting tw
8	38	63.3	226	2 T11145	H+-transporting tw
9	38	63.3	226	2 T11055	H+-transporting tw
10	38	63.3	226	2 S04752	H+-transporting tw
11	38	63.3	293	2 D90153	cobalamin biosynth
12	38	63.3	505	2 H95946	phosphate uptake A
13	38	63.3	561	2 A11409	potassium-transport
14	38	63.3	561	2 A11785	potassium-transport
15	38	63.3	1151	2 T18297	zinc-finger protei
16	37	61.7	166	2 T37052	H+-transporting tw
17	37	61.7	402	2 F90035	TcAB protein limpo
18	37	61.7	484	2 G95293	probable ThrC2 thr
19	36	60.0	220	2 S78395	ribosomal protein
20	36	60.0	403	2 S02856	site-specific DNA-
21	36	60.0	540	2 A75250	carboxylesterase,
22	36	60.0	598	2 A72124	phosphomannomutase
23	36	60.0	598	2 A86498	phosphomannomutase
24	36	60.0	1166	2 C96945	cation efflux syst
25	35	58.3	226	2 T11433	H+-transporting tw
26	35	58.3	226	2 B25188	H+-transporting tw
27	35	58.3	307	2 T48166	hypothetical prote
28	35	58.3	357	2 A86658	oxidoreductase ycg
29	35	58.3	394	2 H72693	hypothetical prote

probable transposase
hypothetical prote
F25A4.24 imported
beta-fructofuranos
photomorphogenesis
semaphorin I precu
hypothetical prote
hypothetical prote
H+-transporting tw
H+-transporting tw
H+-transporting tw
H+-transporting tw
H+-transporting tw
H+-transporting tw
H+-transporting tw

ALIGNMENTS

RESULT 1

PWH06
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 6 - human mitochondrion
N/Alternate names: hydrogen ion-transporting ATP synthase protein 6
C/Species: mitochondrion Homo sapiens (man)
C/Date: 22-May-1981 #sequence_revision 23-Oct-1981 #text_change 09-Jul-2004
C/Accession: A01049
R/Anderson, S.; Bankier, A.T.; Barrell, B.G.; de Bruijn, M.H.L.; Coulson, A.R.; Drouin, S.

Nature 290, 457-465, 1981
A/Title: Sequence and organization of the human mitochondrial genome.
A/Reference number: A00151; MUID:81173052; PMID:7219534
A/Accession: A01049
A/Molecule type: DNA
A/Residues: 1-226 <AND>
A/Cross-references: UNIPROT:P00846; GB:J01415; GB:M12548; GB:M58503; GB:M63932; GB:M63933
C/Genetics:
A/Genes: GDB:MTATP6
A/Cross-references: GDB:118897; OMIM:516060
A/Map position: MTH8527-9207
A/Genome: mitochondrion
A/Genetic code: SGCl
C/Superfamily: H+-transporting ATP synthase protein 6
C/Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxid

Query Match 70.0%; Score 42; DB 1; Length 226;
Best Local Similarity 90.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
| | | | | | | | | |
DB 2 NGNLFASFIA 11

RESULT 2

F59153
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 6 - western lowland gorilla mitochondrion
N/Alternate names: hydrogen ion-transporting ATP synthase protein 6
C/Species: mitochondrion Gorilla gorilla gorilla (western lowland gorilla)
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: F59153
R/Xu, X.; Arnason, U.
Mol. Biol. Evol. 13, 691-698, 1996
A/Title: A complete sequence of the mitochondrial genome of the Western lowland gorilla.
A/Reference number: Z17269; MUID:96212991; PMID:8676744
A/Accession: F59153
A/Status: preliminary; nucleic acid sequence not shown; translation not shown; translated
A/Molecule type: DNA
A/Residues: 1-226 <XUX>
A/Cross-references: UNIPROT:Q9T9Y7; GB:X93347; NID:q1304307; GSPDB:GN00106
A/Note: submitted to GenBank, November 1995
A/Note: this translation is not annotated in GenBank entry GGMITG, release 114.0

C:Genetics:

A:Gene: ATPase6
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 C:Superfamily: H+-transporting ATP synthase protein 6
 C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxid

Query Match 70.0%; Score 42; DB 2; Length 226;
 Best Local Similarity 90.0%; Pred. No. 2;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
 |||||
 DB 2 NENLFASFIA 11
 |||||

RESULT 3
 T10977
 H+-transporting two-sector ATPase (EC 3.6.3.14) protein 6 - pig mitochondrion
 C:Species: mitochondrion Sus scrofa domestica (domestic pig)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
 C:Accession: T10977; T11875; A45954
 R:Lin, C.S.; Liu, C.Y.; Sun, Y.L.; Chang, L.C.; Cheng, I.C.; Yang, P.C.; Mao, S.J.T.; Hu
 submitted to the EMBL Data Library, November 1997
 A:Description: Complete nucleotide sequence of the porcine mitochondrial genome.
 A:Reference number: 217237
 A:Accession: T10977
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-226 <LIN>
 A:Cross-references: UNIPROT:Q35915; EMBL:AF034253; NID:G4958951; PIDN:AA03

R:Ursing, B.M.
 submitted to the EMBL Data Library, February 1999
 A:Description: The complete mitochondrial DNA sequence of the pig (Sus scrofa).
 A:Reference number: 217370
 A:Accession: T11875
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-226 <URS>
 A:Cross-references: EMBL:AJ002189; PIDN:CAA05234.1
 R:Watanabe, T.; Hayashi, Y.; Kimura, J.; Yasuda, Y.; Saitou, N.; Tomita, T.; Ogasawara,
 Biochem. Genet. 24, 385-396, 1986
 A:Title: Pig mitochondrial DNA: polymorphism, restriction map orientation, and sequence
 A:Reference number: A45954; MUID:86295557; PMID:3017295
 A:Accession: A45954
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-28;219-226 <WAT>
 A:Cross-references: GB:M26139
 A:Note: the authors translated the codon AAC for residue 2 as Thr
 C:Genetics:
 A:Gene: ATPase6
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 C:Superfamily: H+-transporting ATP synthase protein 6
 C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxid

Query Match 70.0%; Score 42; DB 2; Length 226;
 Best Local Similarity 90.0%; Pred. No. 2;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
 |||||
 DB 2 NENLFASFIA 11
 |||||

RESULT 4

T11498
 H+-transporting two-sector ATPase (EC 3.6.3.14) protein 6 - dog mitochondrion
 C:Species: mitochondrion Canis lupus familiaris (dog)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
 C:Accession: T11498
 R:Kim, K.S.; Lee, S.E.; Jeong, H.W.; Ha, J.H.

Mol. Phylogenet. Evol. 10, 210-220, 1998
 A:Title: The complete nucleotide sequence of the domestic dog (Canis familiaris) mitochor
 A:Reference number: Z17276; MUID:99097286; PMID:9878232
 A:Accession: T11498
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBI
 A:Molecule type: DNA
 A:Residues: 1-226 <KIM>
 A:Cross-references: UNIPROT:Q9ZZ62; EMBL:U96639; NID:G4154170; PID:G4154176; PIDN:AA00476
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 C:Superfamily: H+-transporting ATP synthase protein 6
 C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxid

Query Match 66.7%; Score 40; DB 2; Length 226;
 Best Local Similarity 75.0%; Pred. No. 4.8;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NGNLFASPIADS 12
 |||||
 DB 2 NENLFASPAAPS 13
 |||||

RESULT 5

JC4363
 Glucagon receptor precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 08-Jan-1996 #sequence_revision 08-Feb-1996 #text_change 09-Jul-2004
 C:Accession: JC4363
 R:Burcelin, R.; Li, J.; Charron, M.J.
 Gene 164, 305-310, 1995
 A:Title: Cloning and sequence analysis of the murine glucagon receptor-encoding gene.
 A:Reference number: JC4363; MUID:96069600; PMID:7590348
 A:Accession: JC4363
 A:Molecule type: mRNA
 A:Residues: 1-485 <BUR>
 A:Cross-references: UNIPROT:Q61606; GB:L38612; NID:G7274387; PIDN:AAF44749.1; PID:G727438
 C:Comment: This receptor is a plasma membrane glycoprotein that belongs to a subfamily of
 glycolysis and gluconeogenesis in li ver and insulin secretion by beta cells.

C:Genetics:
 A:Gene: gr
 A:Introns: 21/3; 55/1; 91/1; 132/3; 168/2; 220/3; 273/1; 294/2; 317/3; 347/2; 393/3; 407/
 C:Superfamily: glucagon receptor
 C:Keywords: glycoprotein; receptor; transmembrane protein
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-485/Product: glucagon receptor #status predicted <MAT>
 F:145-168/Domain: transmembrane #status predicted <TM1>
 F:176-196/Domain: transmembrane #status predicted <TM2>
 F:225-239/Domain: transmembrane #status predicted <TM3>
 F:247-258/Domain: transmembrane #status predicted <TM4>
 F:266-288/Domain: transmembrane #status predicted <TM5>
 F:304-331/Domain: transmembrane #status predicted <TM6>
 F:354-369/Domain: transmembrane #status predicted <TM7>
 F:386-405/Domain: transmembrane #status predicted <TM8>
 F:47.60.65.75.79.118/Binding site: carbohydrate (Asn) #status predicted

Query Match 66.7%; Score 40; DB 2; Length 485;
 Best Local Similarity 77.8%; Pred. No. 11;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NGNLFASFI 9
 |||||
 DB 178 HGNLFASFV 186
 |||||

RESULT 6

JQ1957
 Glucagon receptor - rat
 N:Alternate names: GLP-1 receptor homolog, hepatic
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
 C:Accession: JQ1957; A46211; S29689
 R:Svoboda, M.; Ciccarelli, E.; Tatenoy, M.; Cauvin, A.; Stievenart, M.; Charistophe, J.

Biochem. Biophys. Res. Commun. 191, 479-486, 1993
A:Title: Small introns in a hepatic cDNA encoding a new glucagon-like peptide 1-type red
A:Reference number: JQ1957; MUID:93213282; PMID:8384842
A:Accession: JQ1957
A:Molecule type: mRNA
A:Residues: 1-485 <SVQ>
A:Cross-references: UNIPROT:P30082; GB:X68692
A:Experimental source: liver
R:Jellinek, L.J.; Lok, S.; Rosenberg, G.B.; Smith, R.A.; Grant, F.J.; Biggs, S.; Bensch, Rnan, P.A.; Kindsvogel, W.
Science 259, 1614-1616, 1993
A:Title: Expression cloning and signaling properties of the rat glucagon receptor.
A:Reference number: A46211; MUID:93206036; PMID:8384375
A:Accession: A46211
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 1-37,'W',38-215,'W',217-323,'V',325-485 <JEL>
A:Experimental source: liver
A:Note: sequence extracted from NCBI backbone (NCBIP:127785)
C:Genetics:
C:Introns: 132/3; 168/2; 220/3; 407/3
C:Superfamily: glucagon receptor
C:Keywords: G protein-coupled receptor; glycoprotein; liver; phosphoprotein; transmembra
F:144-167/Domain: transmembrane #status predicted <TM1>
F:175-194/Domain: transmembrane #status predicted <TM2>
F:227-251/Domain: transmembrane #status predicted <TM3>
F:264-286/Domain: transmembrane #status predicted <TM4>
F:305-324/Domain: transmembrane #status predicted <TM5>
F:352-370/Domain: transmembrane #status predicted <TM6>
F:389-405/Domain: transmembrane #status predicted <TM7>
F:47,60,75/Binding site: carboxylate (Asn) (covalent) #status predicted
F:432/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 66.7%; Score 40; DB 2; Length 485;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NGNLFASFI 9
Db 178 NGNLFASFI 186

RESULT 7
PWMS6
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 6 - mouse mitochondrion
N:Alternate names: hydrogen ion-transporting ATP synthase protein 6
C:Species: mitochondrion Mus musculus (house mouse)
C:Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 09-Jul-2004
C:Accession: A01051
R:Bibb, M.J.; Van Etten, R.A.; Wright, C.T.; Walberg, M.W.; Clayton, D.A.
Cell 26, 167-180, 1981
A:Title: Sequence and gene organization of mouse mitochondrial DNA.
A:Reference number: A00153; MUID:82137051; PMID:7332926
A:Accession: A01051
A:Molecule type: DNA
A:Residues: 1-226 <HIB>
A:Cross-references: UNIPROT:P00848; GB:J01420; NID:G342520; PIDN:AAB48649.1; PID:G342523
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: H+-transporting ATP synthase protein 6
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; ox

Query Match 63.3%; Score 38; DB 1; Length 226;
Best Local Similarity 88.9%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFI 9
Db 2 NENLFASFI 10

RESULT 8
PWMS6
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 6 - rat mitochondrion
C:Species: mitochondrion Rattus norvegicus (Norway rat)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: S04752
R:Gadaleta, G.; Pepe, G.; De Candia, G.; Quagliariello, C.; Sbisa, E.; Saccone, C.
J. Mol. Evol. 28, 497-516, 1989
A:Title: The complete nucleotide sequence of the Rattus norvegicus mitochondrial genome:
A:Reference number: S04747; MUID:89362487; PMID:2504926
A:Accession: S04752
A:Molecule type: DNA

T11145
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 6 - fruit bat (Artibeus jamaicens
C:Species: mitochondrion Artibeus jamaicens
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T11145
R:Puno, D.E.; Pinamore, P.S.; Franek, W.R.; Phillips, C.J.; Tarzami, S.; Balzarano, D.
J. Mol. Evol. 47, 703-717, 1998
A:Title: Complete mitochondrial genome of a neotropical fruit bat, Artibeus jamaicensis,
A:Reference number: Z17251; MUID:99065764; PMID:9847413
A:Accession: T11145
A>Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 1-226 <PUM>
A:Cross-references: UNIPROT:O99599; EMBL:AF061340; NID:G4164474; PID:G4164480; PIDN:AAD0
C:Genetics:
A:Genome: mitochondrion
A:Note: ATPase6
C:Superfamily: H+-transporting ATP synthase protein 6
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; ox

Query Match 63.3%; Score 38; DB 2; Length 226;
Best Local Similarity 88.9%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFI 9
Db 2 NENLFASFI 10

RESULT 9
T11055
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 6 - sheep mitochondrion
C:Species: mitochondrion Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T11055
R:Hiendleder, S.; Lewalski, H.; Wassmuth, R.; Janke, A.
J. Mol. Evol. 47, 441-448, 1998
A:Title: The complete mitochondrial DNA sequence of the domestic sheep (Ovis aries) and
A:Reference number: Z17245; MUID:98440761; PMID:9767689
A:Accession: T11055
A>Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 1-226 <HIB>
A:Cross-references: UNIPROT:O78752; EMBL:AF010406; NID:G3445513; PID:G3366628; PIDN:AAD1
C:Experimental source: strain Merinolandschaf; liver
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
A:Note: ATPase6
C:Superfamily: H+-transporting ATP synthase protein 6
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; ox

Query Match 63.3%; Score 38; DB 2; Length 226;
Best Local Similarity 88.9%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFI 9
Db 2 NENLFASFI 10

RESULT 10
S04752
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 6 - rat mitochondrion
C:Species: mitochondrion Rattus norvegicus (Norway rat)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: S04752
R:Gadaleta, G.; Pepe, G.; De Candia, G.; Quagliariello, C.; Sbisa, E.; Saccone, C.
J. Mol. Evol. 28, 497-516, 1989
A:Title: The complete nucleotide sequence of the Rattus norvegicus mitochondrial genome:
A:Reference number: S04747; MUID:89362487; PMID:2504926
A:Accession: S04752
A:Molecule type: DNA

```

A:Residues: 1-226 <GAD>
A:Cross-references: UNIPROT:P05504; EMBL:X14848; NID:g854269; PIDN:CAA32959.1; PID:g1348
C:Genetics:
A:Genome: mitochondrion
A:Superfamily: H+-transporting ATP synthase protein 6
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxid
Query Match 63.3%; Score 38; DB 2; Length 226;
Best Local Similarity 88.9%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 NGNLFASF1 9
DB 2 NENLFASF1 10
RESULT 11
cobalamins biosynthesis protein G (cbiG) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 16-Aug-2004
C:Accession: D90153
R:Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.
arret, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: D90153
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-293 <KUR>
A:Cross-references: UNIPROT:Q980Y9; GB:A506641; NID:g13813263; PIDN:AAK40483.1; GSPDB:G
C:Genetics:
A:Gene: cbiG
C:Superfamily: Cobalamin biosynthesis protein CbiG
Query Match 63.3%; Score 38; DB 2; Length 293;
Best Local Similarity 63.6%; Pred. No. 15;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 GNLFASF1AD 12
DB 80 GSLISFLAD 90
RESULT 12
Phosphate uptake ABC transporter permease protein phoT [imported] - Sinorhizobium meliloti
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C:Accession: H95946
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan-
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A>Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo-
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: H95946
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-505 <KUR>
A:Cross-references: UNIPROT:Q52909; GB:AL591985; PIDN:CAC49240.1; PID:g15140726; GSPDB:G
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: phoT; Smb21174
A:Genome: plasmid
Query Match 63.3%; Score 38; DB 2; Length 505;
Best Local Similarity 88.9%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GNLFASF1A 10
DB 319 GTLFASF1A 327
RESULT 13
potassium-transporting atpase a chain homolog kdpA [imported] - Listeria monocytogenes (s
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: A11409
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A>Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11409
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <GLA>
A:Cross-references: UNIPROT:Q8Y3Z6; GB:NC_003210; PIDN:CAD00895.1; PID:g16412182; GSPDB:G
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: kdpA
C:Superfamily: H+/K+-exchanging ATPase subunit A
Query Match 63.3%; Score 38; DB 2; Length 561;
Best Local Similarity 58.3%; Pred. No. 30;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 NGNLFASF1AD 12
DB 468 NSAFAGFADT 479
RESULT 14
potassium-transporting atpase a chain homolog kdpA [imported] - Listeria innocua (strain
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: A11785
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A>Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11785
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <GLA>
A:Cross-references: UNIPROT:Q927F9; GB:AL592022; PIDN:CAC98056.1; PID:g16415365; GSPDB:G
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: kdpA
C:Superfamily: H+/K+-exchanging ATPase subunit A
Query Match 63.3%; Score 38; DB 2; Length 561;
Best Local Similarity 58.3%; Pred. No. 30;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 NGNLFASFIA 12
 Db 468 NGSAPAGFAADT 479

RESULT 15

T18297
 zinc-finger protein FOG-2 - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C;Accession: T18297
 R;Tevosian, S.G.; Deconinck, A.E.; Cantor, A.B.; Rieff, H.I.; Fujiwara, Y.; Corfas, G.;
 Proc. Natl. Acad. Sci. U.S.A. 96, 950-955, 1999
 A;Title: FOG-2: A novel GATA-family cofactor related to multitype zinc-finger proteins F
 A;Reference number: Z18866; MUID:99128343; PMID:9927674
 A;Accession: T18297
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 1-1151 <REV>
 A;Cross-references: UNIPROT:Q8CCH7; EMBL:AF107306; NID:g4206772; PID:g4206773; PIDN:AAAD1
 C;Genetics:
 A;Gene: FOG-2

Query Match 63.3%; Score 38; DB 2; Length 1151;
 Best Local Similarity 70.0%; Pred. No. 63;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
 Db 928 NGNLFSSHLA 937

Search completed: November 10, 2004, 13:40:43
 Job time : 3.67358 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 14.6528 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-35

Perfect score: 60

Sequence: 1 NGNLFASFIAADS 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	44	73.3	754	Q871Y8	Q871Y8 vibrio para
2	42	70.0	22	Q7J5U6	Q7J5U6 sus scrofa
3	42	70.0	22	AAD05065	AAD05065 sus scrof
4	42	70.0	226	1 ATP6 HUMAN	P00846 homo sapien
5	42	70.0	226	1 ATP6 PIG	Q35915 sus scrofa
6	42	70.0	226	2 Q6R0Z3	Q6R0Z3 homo sapien
7	42	70.0	226	2 Q6R0Z3	Q6R0Z3 homo sapien
8	42	70.0	226	2 Q6R0Z3	Q6R0Z3 homo sapien
9	42	70.0	226	2 Q6RLN5	Q6RLN5 homo sapien
10	42	70.0	226	2 Q6RLU8	Q6RLU8 homo sapien
11	42	70.0	226	2 Q6RMK9	Q6RMK9 homo sapien
12	42	70.0	226	2 Q6RNG3	Q6RNG3 homo sapien
13	42	70.0	226	2 Q6RNI4	Q6RNI4 homo sapien
14	42	70.0	226	2 Q6RBP7	Q6RBP7 homo sapien
15	42	70.0	226	2 Q6RQ23	Q6RQ23 homo sapien
16	42	70.0	226	2 Q6RQ23	Q6RQ23 homo sapien
17	42	70.0	226	2 Q6RQ23	Q6RQ23 homo sapien
18	42	70.0	226	2 Q6RQW5	Q6RQW5 homo sapien
19	42	70.0	226	2 Q6RQW5	Q6RQW5 homo sapien
20	42	70.0	226	2 Q6RQW5	Q6RQW5 homo sapien
21	42	70.0	226	2 Q6RQW5	Q6RQW5 homo sapien
22	42	70.0	226	2 Q6RQW5	Q6RQW5 homo sapien
23	42	70.0	226	2 Q6VHA0	Q6VHA0 homo sapien
24	42	70.0	226	2 Q6VHB0	Q6VHB0 homo sapien
25	42	70.0	226	2 Q6VHC0	Q6VHC0 homo sapien
26	42	70.0	226	2 Q6VHL0	Q6VHL0 homo sapien
27	42	70.0	226	2 Q6VHM2	Q6VHM2 homo sapien
28	42	70.0	226	2 Q6VIX0	Q6VIX0 homo sapien
29	42	70.0	226	2 Q6VJ77	Q6VJ77 homo sapien
30	42	70.0	226	2 Q6VJ13	Q6VJ13 homo sapien
31	42	70.0	226	2 Q6VKS1	Q6VKS1 homo sapien

32 42 70.0 226 2 Q6VL11 Q6VL11 homo sapien
33 42 70.0 226 2 Q6VLP4 Q6VLP4 homo sapien
34 42 70.0 226 2 Q6VLT0 Q6VLT0 homo sapien
35 42 70.0 226 2 Q6WQF4 Q6WQF4 homo sapien
36 42 70.0 226 2 Q71JZ7 Q71JZ7 sus scrofa
37 42 70.0 226 2 Q71K48 Q71K48 sus scrofa
38 42 70.0 226 2 Q7Y6X6 Q7Y6X6 homo sapien
39 42 70.0 226 2 Q7Y788 Q7Y788 homo sapien
40 42 70.0 226 2 Q7Y802 Q7Y802 homo sapien
41 42 70.0 226 2 Q7Y843 Q7Y843 homo sapien
42 42 70.0 226 2 Q7YCC1 Q7YCC1 homo sapien
43 42 70.0 226 2 Q7YCD1 Q7YCD1 homo sapien
44 42 70.0 226 2 Q7YCF8 Q7YCF8 homo sapien
45 42 70.0 226 2 Q7YCG1 Q7YCG1 homo sapien

ALIGNMENTS

RESULT 1
Q871Y8 PRELIMINARY; PRT; 754 AA.
AC Q871Y8
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Prolyl endopeptidase.
GN OrderedLocusNames=VPA0468;
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Nijima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
DR EMBL; AP005085; BAC61811.1; -.
DR HSSP; P23687; 106G.
DR MEROPS; S09.001; -.
DR GO; GO:0003624; F:catalytic activity; IEA.
DR GO; GO:0004287; F:prolyl oligopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002470; Peptidase_S9A.
DR InterPro; IPR004106; Peptidase_S9A.
DR InterPro; IPR002471; Pept_S9_AS.
DR InterPro; IPR00379; Ser esters.
DR Pfam; PF00326; Peptidase_S9; 1.
DR Pfam; PF02897; Peptidase_S9_N; 1.
DR PRINTS; PR00862; PROLOGPTASE.
DR PROSITE; PS00708; PRO-ENDOPEP_SER; 1.
KW Complete proteome.
SQ SEQUENCE 754 AA; 84245 MW; 65EF726912510A02 CRC64;

Query Match 73.3%; Score 44; DB 2; Length 754;

Best Local Similarity 72.7%; Pred. No. 20;

Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GNLFASFIAADS 12

|||||:

Db 402 GNLFASYIVDA 412

RESULT 2

Q7J5U6

ID Q7J5U6 PRELIMINARY; PRT; 22 AA.

AC Q7J5U6;

```

DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE ATPase subunit 6 (Fragment).
OS Sus scrofa (Pig).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=81173052; PubMed=7219534;
RA Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L.,
RA Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A.,
RA Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G.;
RT "Sequence and organization of the human mitochondrial genome.";
RL Nature 290:457-465(1981).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=95132634; PubMed=7530363;
RA Horai S., Hayasaka K., Kondo R., Tsugane K., Takahata N.;
RA "Recent African origin of modern humans revealed by complete sequences
RT of hominoid mitochondrial DNAs.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:532-536(1995).
RN [3]
RN SEQUENCE FROM N.A., AND VARIANT TYR-90.
RP MEDLINE=22062553; PubMed=12022039;
RA Silva W.A. Jr., Bonatto S.L., Holanda A.J., Ribeiro-Dos-Santos A.K.,
RA Paixao B.M., Goldman G.H., Abe-Sandes K., Rodriguez-Delfin L.,
RA Barbosa M., Peco-Larson M.L., Petzl-Erler M.L., Valente V.,
RA Santos S.E., Zago M.A.;
RT "Mitochondrial genome diversity of native Americans supports a single
RT early entry of founder populations into America.";
RL Am. J. Hum. Genet. 71:187-192(2002).
RN [4]
RN VARIANTS ALA-59 AND ILE-213.
RX MEDLINE=89072713; PubMed=3201231;
RA Wallace D.C., Singh G., Lott M.T., Hodge J.A., Schurr T.G.,
RA Lezza A.M., Eissas L.J. II, Nikoskelainen E.K.;
RA "Mitochondrial DNA mutation associated with Leber's hereditary optic
RT neuropathy.";
RL Science 242:1427-1430(1988).
RN [5]
RN VARIANT NARP ARG-156.
RX MEDLINE=90178075; PubMed=2137962;
RA Holt I.J., Harding A.E., Petty R.K., Morgan-Hughes J.A.;
RA "A new mitochondrial disease associated with mitochondrial DNA
RT heteroplasmy.";
RL Am. J. Hum. Genet. 46:428-433(1990).
RN [6]
RN VARIANTS SER-33; ALA-59; TYR-61; TYR-90; ALA-112; THR-155; ILE-213 AND
RP GLY-219.
RX MEDLINE=92098084; PubMed=1757091;
RA Marzuki S., Noer A.S., Lertit P., Thyagarajan D., Kapsa R.,
RA Utthanaphol P., Byrne E.;
RT "Normal variants of human mitochondrial DNA and translation products:
RT the building of a reference data base.";
RL Hum. Genet. 88:139-145(1991).
RN [7]
RN VARIANT LS PRO-156.
RX MEDLINE=93371018; PubMed=8395787;
RA de Vries D.D., van Engelen B.G.M., Gabreels F.J.M., Ruitenbeek W.,
RA van Oost B.A.;
RT "A second missense mutation in the mitochondrial ATPase 6 gene in
RT Leigh's syndrome.";
RL Ann. Neurol. 34:410-412(1993).
RN [8]
RN VARIANT LHON THR-192.
RX MEDLINE=95243243; PubMed=7726182;
RA Lamminen T., Majander A., Juvenon V., Wikstrom M., Aula P.,
RA Nikoskelainen E., Savontaus M.-L.;
RT "A mitochondrial mutation at nt 9101 in the ATP synthase 6 gene
RT associated with deficient oxidative phosphorylation in a family with
RT Leber hereditary optic neuropathy.";

```

```

DE ATP synthase a chain (EC 3.6.3.14) (ATPase protein 6).
GN Name=MTATP6; Synonyms=ATP6;
OG Homo sapiens (Human).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=81173052; PubMed=7219534;
RA Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L.,
RA Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A.,
RA Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G.;
RT "Sequence and organization of the human mitochondrial genome.";
RL Nature 290:457-465(1981).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=95132634; PubMed=7530363;
RA Horai S., Hayasaka K., Kondo R., Tsugane K., Takahata N.;
RA "Recent African origin of modern humans revealed by complete sequences
RT of hominoid mitochondrial DNAs.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:532-536(1995).
RN [3]
RN SEQUENCE FROM N.A., AND VARIANT TYR-90.
RP MEDLINE=22062553; PubMed=12022039;
RA Silva W.A. Jr., Bonatto S.L., Holanda A.J., Ribeiro-Dos-Santos A.K.,
RA Paixao B.M., Goldman G.H., Abe-Sandes K., Rodriguez-Delfin L.,
RA Barbosa M., Peco-Larson M.L., Petzl-Erler M.L., Valente V.,
RA Santos S.E., Zago M.A.;
RT "Mitochondrial genome diversity of native Americans supports a single
RT early entry of founder populations into America.";
RL Am. J. Hum. Genet. 71:187-192(2002).
RN [4]
RN VARIANTS ALA-59 AND ILE-213.
RX MEDLINE=89072713; PubMed=3201231;
RA Wallace D.C., Singh G., Lott M.T., Hodge J.A., Schurr T.G.,
RA Lezza A.M., Eissas L.J. II, Nikoskelainen E.K.;
RA "Mitochondrial DNA mutation associated with Leber's hereditary optic
RT neuropathy.";
RL Science 242:1427-1430(1988).
RN [5]
RN VARIANT NARP ARG-156.
RX MEDLINE=90178075; PubMed=2137962;
RA Holt I.J., Harding A.E., Petty R.K., Morgan-Hughes J.A.;
RA "A new mitochondrial disease associated with mitochondrial DNA
RT heteroplasmy.";
RL Am. J. Hum. Genet. 46:428-433(1990).
RN [6]
RN VARIANTS SER-33; ALA-59; TYR-61; TYR-90; ALA-112; THR-155; ILE-213 AND
RP GLY-219.
RX MEDLINE=92098084; PubMed=1757091;
RA Marzuki S., Noer A.S., Lertit P., Thyagarajan D., Kapsa R.,
RA Utthanaphol P., Byrne E.;
RT "Normal variants of human mitochondrial DNA and translation products:
RT the building of a reference data base.";
RL Hum. Genet. 88:139-145(1991).
RN [7]
RN VARIANT LS PRO-156.
RX MEDLINE=93371018; PubMed=8395787;
RA de Vries D.D., van Engelen B.G.M., Gabreels F.J.M., Ruitenbeek W.,
RA van Oost B.A.;
RT "A second missense mutation in the mitochondrial ATPase 6 gene in
RT Leigh's syndrome.";
RL Ann. Neurol. 34:410-412(1993).
RN [8]
RN VARIANT LHON THR-192.
RX MEDLINE=95243243; PubMed=7726182;
RA Lamminen T., Majander A., Juvenon V., Wikstrom M., Aula P.,
RA Nikoskelainen E., Savontaus M.-L.;
RT "A mitochondrial mutation at nt 9101 in the ATP synthase 6 gene
RT associated with deficient oxidative phosphorylation in a family with
RT Leber hereditary optic neuropathy.";

```

```

DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE ATPase subunit 6 (Fragment).
OS Sus scrofa (Pig).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=81173052; PubMed=7219534;
RA Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L.,
RA Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A.,
RA Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G.;
RT "Sequence and organization of the human mitochondrial genome.";
RL Nature 290:457-465(1981).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Peripheral blood;
RC MEDLINE=98403943; PubMed=9734874;
RA Tartaglia M., Saulle E.;
RT "Rapid communication: nucleotide sequence of porcine and ovine
RT tRNA(Lys) and ATPase8 mitochondrial genes.";
RL J. Anim. Sci. 76:2207-2208(1998).
DR EMBL; AF039170; AAD05065.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
KW Mitochondrion.
FT NON TER 22 22
SQ SEQUENCE 22 AA; 2411 MW; D440C56CAE0EALC9 CRC64;

Query Match 70.0%; Score 42; DB 2; Length 22;
Best Local Similarity 90.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLPASPIA 10
DB 2 NENLPASPIA 11

RESULT 3
AAD05065 PRELIMINARY; PRT; 22 AA.
AC AAD05065;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE ATPase subunit 6 (Fragment).
OS Sus scrofa (Pig).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=81173052; PubMed=7219534;
RA Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L.,
RA Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A.,
RA Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G.;
RT "Sequence and organization of the human mitochondrial genome.";
RL Nature 290:457-465(1981).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Peripheral blood;
RC MEDLINE=98403943; PubMed=9734874;
RA Tartaglia M., Saulle E.;
RT "Rapid communication: nucleotide sequence of porcine and ovine
RT tRNA(Lys) and ATPase8 mitochondrial genes.";
RL J. Anim. Sci. 76:2207-2208(1998).
DR EMBL; AF039170; AAD05065.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
KW Mitochondrion.
FT NON TER 22 22
SQ SEQUENCE 22 AA; 2411 MW; D440C56CAE0EALC9 CRC64;

Query Match 70.0%; Score 42; DB 2; Length 22;
Best Local Similarity 90.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLPASPIA 10
DB 2 NENLPASPIA 11

RESULT 4
ATP6 HUMAN STANDARD; PRT; 226 AA.
AC P08436; Q34772; Q8HNO4;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)

```

Am. J. Hum. Genet. 56:1238-1240 (1995).
[9] VARIANT FBSN PRO-217.
RX MEDLINE=95398422; PubMed=7668837;
RA Thyagarajan D., Shanske S., Vazquez-Memije M., de Vivo D., Dimauro S.;
RT "A novel mitochondrial ATPase 6 point mutation in familial bilateral
RT striatal necrosis";
RL Ann. Neurol. 38:468-472 (1995).
[10] VARIANT LS PRO-217.
RX MEDLINE=97416654; PubMed=9270604;
RA Campos Y., Martin M.A., Rubio J.C., Solana L.G., Garcia-Benayas C.,
RA Terradas J.L., Arenas J.;
RT "Leigh syndrome associated with the T9176C mutation in the ATPase 6
RT gene of mitochondrial DNA";
RL Neurology 49:595-597 (1997).
[11] VARIANT LS ARG-156.
RX MEDLINE=98220421; PubMed=9556461;
RA Takahashi S., Makita Y., Oki J., Miyamoto A., Yanagawa J., Naito E.,
RA Goto Y., Okuno A.;
RT "De novo mtDNA nt 8993 (T-G) mutation resulting in Leigh syndrome.";
RL Am. J. Hum. Genet. 62:717-719 (1998).
[12] VARIANT LS PRO-217.
RX MEDLINE=98161883; PubMed=9501263;
RA Dionisi-Vici C., Seneca S., Zeviani M., Fariello G., Rimoldi M.,
RA Bertini E., De Meirleir L.;
RT "Fulminant Leigh syndrome and sudden unexpected death in a family with
RT the T9176C mutation of the mitochondrial ATPase 6 gene";
RL J. Inher. Metab. Dis. 21:2-8 (1998).
[13] VARIANT THR-155.
RX MEDLINE=98127994; PubMed=9461455;
RA Rieder M.J., Taylor S.L., Tobe V.O., Nickerson D.A.;
RT "Automating the identification of DNA variations using quality-based
RT fluorescence re-sequencing: analysis of the human mitochondrial
RT genome";
RL Nucleic Acids Res. 26:967-973 (1998).
CC -!- FUNCTION: Key component of the proton channel; it may play a
CC direct role in the translocation of protons across the membrane.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC has three main subunits: a, b and c.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -!- DISEASE: Defects in WTATP6 are the cause of neurogenic muscle
CC weakness, ataxia, and retinitis pigmentosa (NARP) [MIM:551500].
CC -!- DISEASE: Defects in WTATP6 are a cause of Leber's hereditary optic
CC neuropathy (LHON) [MIM:535000]; also known as Leber's optic
CC atrophy. LHON is a maternally inherited disease resulting in acute
CC bilateral blindness due to retinal degeneration predominantly in
CC young men. Cardiac conduction defects and neurological defects
CC have also been described.
CC -!- DISEASE: Defects in WTATP6 are a cause of Leigh syndrome (LS)
CC [MIM:256000]. LS is a severe neurological disorder characterized
CC by bilaterally symmetrical necrotic lesions in subcortical brain
CC regions.
CC -!- DISEASE: Defects in WTATP6 may be a cause of familial bilateral
CC striatal necrosis (FBSN). FBSN is a neurological disorder
CC resembling Leigh syndrome.
CC -!- SIMILARITY: Belongs to the ATPase A chain family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: J01415; AB85948.1; --
DR EMBL: V00662; CA24031.1; --
DR EMBL: D38112; BAA07295.1; --
DR EMBL: AF465962; AAN14772.1; --
DR PIR: A01049; PWHU6.
DR Genew: HGNC:7414; WTATP6.
DR Reactome: P00846; --
DR MIM: 516060; --
DR MIM: 551500; --
DR MIM: 535000; --
DR MIM: 256000; --
DR GO: GO:0000276; C:proton-transporting ATP synthase complex, c. .; NAS.
DR GO: GO:0015986; P:ATP synthesis coupled proton transport; NAS.
DR InterPro: IPR000568; ATPsynt_Asub.
DR Pfam: PF00119; ATP-synt_A; 1.
DR PRINTS: PR00123; ATPASEA.
DR TIGRfams: TIGR01131; ATP_synt_6_or_A; 1.
DR PROSITE: PS00449; ATPASEA; 1.
KW CF(0); Disease mutation; Hydrogen ion transport;
KW Leber hereditary optic neuropathy; Leigh syndrome; Mitochondrion;
KW Polymorphism; Transmembrane.
FT VARIAT 33 33 T -> S.
FT VARIAT 59 59 /FTid=VAR_008556.
FT VARIAT 61 61 /FTid=VAR_000792.
FT VARIAT 90 90 H -> Y.
FT VARIAT 90 90 H -> Y.
FT VARIAT 112 112 H -> Y.
FT VARIAT 155 155 T -> A.
FT VARIAT 155 155 /FTid=VAR_008559.
FT VARIAT 156 156 A -> T.
FT VARIAT 156 156 /FTid=VAR_008560.
FT VARIAT 156 156 L -> R (in NARP and LS).
FT VARIAT 156 156 /FTid=VAR_000793.
FT VARIAT 177 177 L -> P (in LS).
FT VARIAT 177 177 /FTid=VAR_000794.
FT VARIAT 192 192 A -> I.
FT VARIAT 192 192 /FTid=VAR_008561.
FT VARIAT 213 213 I -> T (in LHON; possible rate primary
FT VARIAT 213 213 mutation).
FT VARIAT 217 217 /FTid=VAR_000795.
FT VARIAT 217 217 V -> I.
FT VARIAT 217 217 /FTid=VAR_000796.
FT VARIAT 219 219 L -> P (in LS and in familial bilateral
FT VARIAT 219 219 striatal necrosis).
FT VARIAT 219 219 /FTid=VAR_000797.
FT VARIAT 219 219 S -> G.
QY 1 KGNLPASPIA 10 Query Match 70.0%; Score 42; DB 1; Length 226;
Db 2 NENLPASPIA 11 Best Local Similarity 90.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
RESULT 5
ATP6_PIG STANDARD; PRT; 226 AA.
AC Q35915; O79878;
DT 15-DEC-1998 (Rel. 37; Created)
DT 15-DEC-1998 (Rel. 37; Last sequence update)
DT 05-JUL-2004 (Rel. 44; Last annotation update)
DE ATP synthase a chain (EC 3.6.3.14) (ATPase protein 6).
GN Name=WTATP6; Synonyms=ATP6, ATPASE6;
OS Sus scrofa (Pig).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OC NCBI_TaxID=9823;

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=98404150; PubMed=9732457;
RA Ursing B.M., Arnason U.;
RT "The complete mitochondrial DNA sequence of the pig (Sus scrofa).";
RL J. Mol. Evol. 47:302-306(1998).
RN RP SEQUENCE FROM N.A.
RC STRAIN=Landrace; TISSUE=Heart;
RX MEDLINE=99365306; PubMed=10433971;
RA Lin C.S., Sun Y.L., Liu C.Y., Yang P.C., Chang L.C., Cheng I.C.,
RA Mao S.J.T., Huang M.C.;
RT "Complete nucleotide sequence of pig (Sus scrofa) mitochondrial genome
and dating evolutionary divergence within artiodactyla.";
RL Gene 236:107-114(1999).
RN RP SEQUENCE OF 1-37 FROM N.A.
RX MEDLINE=86295557; PubMed=3017295;
RA Watanabe T., Hayashi Y., Kimura J., Yasuda Y., Saitou N., Tomita T.,
RA Ogasawara N.;
RT "Pig mitochondrial DNA: polymorphism, restriction map orientation, and
sequence data.";
RL Biochem. Genet. 24:385-396(1986).
CC -!- FUNCTION: Key component of the proton channel; it may play a
direct role in the translocation of protons across the membrane.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
core - and CF(0) - the membrane proton channel. CF(1) has five
subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
has three main subunits: a, b and c.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
inner membrane.
CC -!- SIMILARITY: Belongs to the ATPase A chain family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR ENBL; AJ002189; CAA05234.1; -.
DR ENBL; AF034253; AAD34190.1; -.
DR ENBL; M26139; AAA32031.1; -.
DR PIR; T10977; T10977.
DR InterPro; IPR000568; ATPsynth_Asub.
DR Pfam; PF00119; ATP-synt_A; 1.
DR PRINTS; PR00123; ATPASEA.
DR TIGRFAMs; TIGR01131; ATP synt 6_or_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1.
DR CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT CONFLICT 30 37 LFTPKEL -> SLVLDNT (in Ref. 3).
SQ SEQUENCE 226 AA; 25039 MW; BE197F3B086FEA2 CRC64;

Query Match 70.0%; Score 42; DB 1; Length 226;
Best Local Similarity 90.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
DB 2 NENLFASFIA 11

RESULT 6
Q6R023 PRELIMINARY; PRT; 226 AA.
AC Q6R023;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE ATP synthase F0 subunit 6.

Query Match 70.0%; Score 42; DB 1; Length 226;
Best Local Similarity 90.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
DB 2 NENLFASFIA 11

RESULT 6
Q6R023 PRELIMINARY; PRT; 226 AA.
AC Q6R023;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE ATP synthase F0 subunit 6.

```

```

GN Name=ATP6;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN RP SEQUENCE FROM N.A.
RA Starikovskaya E.B., Sukernik R.I., Derbeneva O.A., Volodko N.V.,
RA Tortoni A., Ruiz-Pesini E., Brown M.D., Lott M.T., Hosseini S.H.,
RA Huoponen K., Wallace D.C.;
RT "Mitochondrial DNA Diversity in Indigenous Populations of Southern
RT Extent of Siberia, and the Origins of Native American Haplogroups.";
RL Ann. Hum. Genet. 0:0-0(2004).
CC -!- FUNCTION: Key component of the proton channel; it may play a
direct role in the translocation of protons across the membrane
(BY similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
core - and CF(0) - the membrane proton channel. CF(1) has five
subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
has three main subunits: a, b and c (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ATPase A chain family.
DR ENBL; AY519486; AAR91237.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR000568; ATPsynth_Asub.
DR Pfam; PF00119; ATP-synt_A; 1.
DR PRINTS; PR00123; ATPASEA.
DR TIGRFAMs; TIGR01131; ATP synt 6_or_A; 1.
DR PROSITE; PS00449; ATPASE_A; UNKNOWN 1.
DR CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
Transmembrane; Transport.
SQ SEQUENCE 226 AA; 24783 MW; 61D5F6712A993712 CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
Best Local Similarity 90.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
DB 2 NENLFASFIA 11

RESULT 7
Q6RXX7 PRELIMINARY; PRT; 226 AA.
AC Q6RXX7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE ATP synthase F0 subunit 6.
GN Name=ATP6;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- FUNCTION: Key component of the proton channel; it may play a
direct role in the translocation of protons across the membrane
(BY similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic

```

CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: a, b and c (By similarity).
 CC - SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC - SIMILARITY: Belongs to the ATPase A chain family.
 DR EMBL; AY495323; AAR95530.1; -
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR000568; ATPsynt_Asub.
 DR Pfam; PF00119; ATP-synt_A; 1.
 DR PRINTS; PR00123; ATPASEA.
 DR TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
 DR PROSITE; PS00449; ATPASE_A; UNKNOW 1.
 KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
 KW Transmembrane; Transport.
 SQ SEQUENCE 226 AA; 24783 MW; 578BEB0432E140FA CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
 Best Local Similarity 90.0%; Pred. No. 14;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASPIA 10
 DB 2 NENLFASPIA 11

RESULT 8
 Q6RKY6 PRELIMINARY; PRT; 226 AA.
 ID Q6RKY6
 AC Q6RKY6
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE ATP synthase F0 subunit 6.
 GN Name=ATP6;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14760490;
 RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
 RA Irwin J.A., Parsons T.J.;
 RT "Single nucleotide polymorphisms over the entire mtDNA genome that
 RT increase the power of forensic testing in Caucasians.";
 RL Int. J. Legal Med. 0:0-0(2004).
 CC -!- FUNCTION: Key component of the proton channel; it may play a
 CC direct role in the translocation of protons across the membrane
 CC (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 CC H(+) (Out).
 CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: a, b and c (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- SIMILARITY: Belongs to the ATPase A chain family.
 DR EMBL; AY495322; AAR95517.1; -
 DR EMBL; AY495317; AAR95452.1; -
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR000568; ATPsynt_Asub.
 DR Pfam; PF00119; ATP-synt_A; 1.
 DR PRINTS; PR00123; ATPASEA.
 DR TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
 DR PROSITE; PS00449; ATPASE_A; 1.
 KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
 KW Transmembrane; Transport.
 SQ SEQUENCE 226 AA; 24797 MW; 578PFF1436F550FA CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
 Best Local Similarity 90.0%; Pred. No. 14;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASPIA 10
 DB 2 NENLFASPIA 11

RESULT 9
 Q6RLN5 PRELIMINARY; PRT; 226 AA.
 ID Q6RLN5
 AC Q6RLN5
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE ATP synthase F0 subunit 6.
 GN Name=ATP6;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14760490;
 RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
 RA Irwin J.A., Parsons T.J.;
 RT "Single nucleotide polymorphisms over the entire mtDNA genome that
 RT increase the power of forensic testing in Caucasians.";
 RL Int. J. Legal Med. 0:0-0(2004).
 CC -!- FUNCTION: Key component of the proton channel; it may play a
 CC direct role in the translocation of protons across the membrane
 CC (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 CC H(+) (Out).
 CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: a, b and c (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- SIMILARITY: Belongs to the ATPase A chain family.
 DR EMBL; AY495295; AAR95166.1; -
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR000568; ATPsynt_Asub.
 DR Pfam; PF00119; ATP-synt_A; 1.
 DR PRINTS; PR00123; ATPASEA.
 DR TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
 DR PROSITE; PS00449; ATPASE_A; 1.
 KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
 KW Transmembrane; Transport.
 SQ SEQUENCE 226 AA; 24775 MW; B59C1EE8DF54F0E CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
 Best Local Similarity 90.0%; Pred. No. 14;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASPIA 10
 DB 2 NENLFASPIA 11

RESULT 10
 Q6RLUS PRELIMINARY; PRT; 226 AA.
 ID Q6RLUS
 AC Q6RLUS
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE ATP synthase F0 subunit 6.
 GN Name=ATP6;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

RN SEQUENCE FROM N.A.
 RX PubMed=14760490;
 RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
 RA Irwin J.A., Parsons T.J.;
 RA "Single nucleotide polymorphisms over the entire mtDNA genome that
 RT increase the power of forensic testing in Caucasians.";
 RL Int. J. Legal Med. 0:0-0(2004).
 CC -!- FUNCTION: Key component of the proton channel; it may play a
 CC direct role in the translocation of protons across the membrane
 CC (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 CC H(+) (Out).
 CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: a, b and c (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- SIMILARITY: Belongs to the ATPase A chain family.
 DR EMBL; AY495288; AAR95075.1; -.
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR000568; ATPsynt_Asub.
 DR Pfam; PF00119; ATP-synt A; 1.
 DR PRINTS; PR00123; ATPASEA.
 DR TIGRFAMs; TIGR01131; ATP synt 6_or_A; 1.
 DR PROSITE; PS00449; ATPASE_A; 1; UNKNOW_1.
 KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
 KW Transmembrane; transport.
 SQ SEQUENCE 226 AA; 24817 MW; 68B0FF0F37455207 CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
 Best Local Similarity 90.0%; Pred. No. 14;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
 DB 2 NENLFASFIA 11

RESULT 11
 Q6RMK9 PRELIMINARY; PRT; 226 AA.
 AC Q6RMK9;
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE ATP synthase F0 subunit 6.
 GN Name=ATP6;
 OS Homo sapiens (Human).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14760490;
 RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
 RA Irwin J.A., Parsons T.J.;
 RA "Single nucleotide polymorphisms over the entire mtDNA genome that
 RT increase the power of forensic testing in Caucasians.";
 RL Int. J. Legal Med. 0:0-0(2004).
 CC -!- FUNCTION: Key component of the proton channel; it may play a
 CC direct role in the translocation of protons across the membrane
 CC (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 CC H(+) (Out).
 CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: a, b and c (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- SIMILARITY: Belongs to the ATPase A chain family.
 DR EMBL; AY495288; AAR94711.1; -.

DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR000568; ATPsynt_Asub.
 DR Pfam; PF00119; ATP-synt A; 1.
 DR PRINTS; PR00123; ATPASEA.
 DR TIGRFAMs; TIGR01131; ATP synt 6_or_A; 1.
 DR PROSITE; PS00449; ATPASE_A; UNKNOW_1.
 KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
 KW Transmembrane; transport.
 SQ SEQUENCE 226 AA; 24787 MW; 68B6236430292207 CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
 Best Local Similarity 90.0%; Pred. No. 14;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
 DB 2 NENLFASFIA 11

RESULT 12
 Q6RNG3 PRELIMINARY; PRT; 226 AA.
 AC Q6RNG3;
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE ATP synthase F0 subunit 6.
 GN Name=ATP6;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14760490;
 RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
 RA Irwin J.A., Parsons T.J.;
 RA "Single nucleotide polymorphisms over the entire mtDNA genome that
 RT increase the power of forensic testing in Caucasians.";
 RL Int. J. Legal Med. 0:0-0(2004).
 CC -!- FUNCTION: Key component of the proton channel; it may play a
 CC direct role in the translocation of protons across the membrane
 CC (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 CC H(+) (Out).
 CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: a, b and c (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- SIMILARITY: Belongs to the ATPase A chain family.
 DR EMBL; AY495229; AAR94308.1; -.
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR000568; ATPsynt_Asub.
 DR Pfam; PF00119; ATP-synt A; 1.
 DR PRINTS; PR00123; ATPASEA.
 DR TIGRFAMs; TIGR01131; ATP synt 6_or_A; 1.
 DR PROSITE; PS00449; ATPASE_A; 1.
 KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
 KW Transmembrane; transport.
 SQ SEQUENCE 226 AA; 24787 MW; DCB98F0136P5520F CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
 Best Local Similarity 90.0%; Pred. No. 14;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLFASFIA 10
 DB 2 NENLFASFIA 11

RESULT 13

```

Q6RNI4
ID Q6RNI4 PRELIMINARY; PRT; 226 AA.
AC Q6RNI4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP synthase F0 subunit 6.
GN Name=ATP6;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- FUNCTION: Key component of the proton channel; it may play a
CC direct role in the translocation of protons across the membrane
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC has three main subunits: a, b and c (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ATPase A chain family.
DR EMBL; AY495227; AAR94282.1; -.
DR EMBL; AY495206; AAR94009.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR000568; ATPsynt_Asub.
DR Pfam; PF00119; ATP-synt_A; 1.
DR PRINTS; PR00123; ATPASEA.
DR TIGRfams; TIGR01131; ATP synt 6_or_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1_-or_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1_-or_A; 1.
KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
KW Transmembrane; Transport.
SQ SEQUENCE 226 AA; 24917 MW; 68B98F0136F55212 CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
Best Local Similarity 90.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NGNLFASPIA 10
Db 2 NENLFASPIA 11
| | | | |
| | | | |

RESULT 14
Q6RPB7 PRELIMINARY; PRT; 226 AA.
ID Q6RPB7;
AC Q6RPB7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP synthase F0 subunit 6.
GN Name=ATP6;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- FUNCTION: Key component of the proton channel; it may play a
CC direct role in the translocation of protons across the membrane
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC has three main subunits: a, b and c (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ATPase A chain family.
DR EMBL; AY495227; AAR94282.1; -.
DR EMBL; AY495206; AAR94009.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR000568; ATPsynt_Asub.
DR Pfam; PF00119; ATP-synt_A; 1.
DR PRINTS; PR00123; ATPASEA.
DR TIGRfams; TIGR01131; ATP synt 6_or_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1_-or_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1_-or_A; 1.
KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
KW Transmembrane; Transport.
SQ SEQUENCE 226 AA; 24917 MW; 68B98F0136F55212 CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
Best Local Similarity 90.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NGNLFASPIA 10
Db 2 NENLFASPIA 11
| | | | |
| | | | |

RESULT 14
Q6RQ23 PRELIMINARY; PRT; 226 AA.
ID Q6RQ23;
AC Q6RQ23;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP synthase F0 subunit 6.
GN Name=ATP6;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- FUNCTION: Key component of the proton channel; it may play a
CC direct role in the translocation of protons across the membrane
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC has three main subunits: a, b and c (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ATPase A chain family.
DR EMBL; AY495171; AAR93554.1; -.
DR EMBL; AY495171; AAR93554.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR000568; ATPsynt_Asub.
DR Pfam; PF00119; ATP-synt_A; 1.
DR PRINTS; PR00123; ATPASEA.
DR TIGRfams; TIGR01131; ATP synt 6_or_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1.

```

```

RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- FUNCTION: Key component of the proton channel; it may play a
CC direct role in the translocation of protons across the membrane
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC has three main subunits: a, b and c (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ATPase A chain family.
DR EMBL; AY495195; AAR93918.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR000568; ATPsynt_Asub.
DR Pfam; PF00119; ATP-synt_A; 1.
DR PRINTS; PR00123; ATPASEA.
DR TIGRfams; TIGR01131; ATP synt 6_or_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1_-or_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1_-or_A; 1.
KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
KW Transmembrane; Transport.
SQ SEQUENCE 226 AA; 24773 MW; 68B0ECAC0D655207 CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
Best Local Similarity 90.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NGNLFASPIA 10
Db 2 NENLFASPIA 11
| | | | |
| | | | |

RESULT 15
Q6RQ23 PRELIMINARY; PRT; 226 AA.
ID Q6RQ23;
AC Q6RQ23;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP synthase F0 subunit 6.
GN Name=ATP6;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- FUNCTION: Key component of the proton channel; it may play a
CC direct role in the translocation of protons across the membrane
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC has three main subunits: a, b and c (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ATPase A chain family.
DR EMBL; AY495171; AAR93554.1; -.
DR EMBL; AY495171; AAR93554.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR000568; ATPsynt_Asub.
DR Pfam; PF00119; ATP-synt_A; 1.
DR PRINTS; PR00123; ATPASEA.
DR TIGRfams; TIGR01131; ATP synt 6_or_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1.

```

KW CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
KW Transmembrane; Transport.
SQ SEQUENCE 226 AA; 24773 MW; 73B0FF12EDF79903 CRC64;

Query Match 70.0%; Score 42; DB 2; Length 226;
Best Local Similarity 90.0%; Pred. NO. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNLPASPIA 10
| | | | | | | | | |
Db 2 NENLPASPIA 11

Search completed: November 10, 2004, 13:38:35
Job time : 15.6528 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 10.0173 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-36

Perfect score: 155

Sequence: 1 ILTSPWTSSGLMPRLQKAAEPKLNQP 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A COMB pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTCUS COMB pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1 pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	32.3	196	4	US-09-252-991A-21163 Sequence 21163, A
2	49	31.6	462	4	US-09-252-991A-18304 Sequence 18304, A
3	48.5	31.3	149	4	US-09-252-991A-20842 Sequence 20842, A
4	48.5	31.3	343	2	US-08-856-444-2 Sequence 2, Appli
5	48.5	31.3	879	1	US-08-554-612C-1 Sequence 1, Appli
6	47.5	30.6	361	4	US-08-583-110-5129 Sequence 5129, Ap
7	47.5	30.6	611	4	US-08-252-991A-29866 Sequence 29866, A
8	47	30.3	39	4	US-09-881-710-12 Sequence 12, Appl
9	47	30.3	39	4	US-09-881-710-30 Sequence 30, Appl
10	47	30.3	40	4	US-09-881-710-28 Sequence 28, Appl
11	47	30.3	76	4	US-09-881-710-1 Sequence 1, Appli
12	47	30.3	84	4	US-09-513-999C-7233 Sequence 7233, Ap
13	47	30.3	146	4	US-08-252-991A-20461 Sequence 20461, A
14	47	30.3	152	4	US-09-187-999-2 Sequence 2, Appli
15	47	30.3	220	4	US-09-252-991A-25179 Sequence 25179, A
16	47	30.3	224	4	US-09-252-991A-21765 Sequence 21765, A
17	47	30.3	281	2	US-08-822-701-9 Sequence 9, Appli
18	47	30.3	281	3	US-08-935-855-9 Sequence 91, Appl
19	47	30.3	281	4	US-09-538-092-91 Sequence 16145, A
20	47	30.3	401	4	US-09-248-796A-16145 Sequence 16145, A
21	47	30.3	658	4	US-09-248-796A-17674 Sequence 4, Appli
22	47	30.3	821	1	US-07-935-311A-4 Sequence 4, Appli
23	47	30.3	821	1	US-08-368-079-4 Sequence 33, Appli
24	47	30.3	821	4	US-09-886-319A-33 Sequence 4, Appli
25	47	30.3	821	5	PCT-US93-07996-4 Sequence 34, Appli
26	47	30.3	822	4	US-09-886-319A-34 Sequence 31583, A
27	46.5	30.0	194	4	US-09-252-991A-31583 Sequence 31583, A

28	46.5	30.0	289	4	US-09-252-991A-27216 Sequence 27216, A
29	46.5	30.0	351	4	US-09-252-991A-23740 Sequence 23740, A
30	46.5	30.0	433	4	US-09-107-532A-5822 Sequence 5822, Ap
31	46	29.7	269	4	US-03-270-767-46698 Sequence 46698, A
32	46	29.7	448	4	US-09-198-452A-370 Sequence 370, App
33	46	29.7	461	4	US-09-252-991A-26815 Sequence 26815, A
34	46	29.7	537	3	US-09-134-078-23 Sequence 23, Appli
35	46	29.7	863	4	US-09-252-991A-19574 Sequence 19574, A
36	45.5	29.4	186	4	US-09-252-991A-24198 Sequence 24198, A
37	45.5	29.4	194	4	US-09-248-796A-23028 Sequence 23028, A
38	45.5	29.4	499	4	US-09-054-272-36 Sequence 36, Appli
39	45.5	29.4	499	4	US-09-054-272-51 Sequence 51, Appli
40	45.5	29.4	505	3	US-09-257-581-5 Sequence 5, Appli
41	45.5	29.4	505	3	US-09-257-581-7 Sequence 7, Appli
42	45.5	29.4	734	4	US-09-602-787A-296 Sequence 296, App
43	45	29.0	109	4	US-09-328-352-4405 Sequence 4405, Ap
44	45	29.0	161	4	US-09-252-991A-30294 Sequence 30294, A
45	45	29.0	227	4	US-09-252-991A-30379 Sequence 30379, A

ALIGNMENTS

RESULT 1
US-09-252-991A-21163
; Sequence 21163, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21163
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21163

Query Match 32.3%; Score 50; DB 4; Length 196;
Best Local Similarity 45.0%; Pred. NO. 11;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 3 TSPWTTSSGLMPRLQKAAEA 22
:||||:||||:|
DB 3 SSNWTSSAGAKSRTSGASTA 22

RESULT 2
US-09-252-991A-18304
; Sequence 18304, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18304
; LENGTH: 462
; TYPE: PRT

```
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18304

Query Match          31.6%; Score 49; DB 4; Length 462;
Best Local Similarity 53.8%; Pred. No. 44;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY  2 LTSPWTTSSGLWP 14
    :|||:||:
DB  24 ISSPWATSAAGWP 36

RESULT 3
US-09-252-991A-20842
; Sequence 20842, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20842
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20842

Query Match          31.3%; Score 48.5; DB 4; Length 149;
Best Local Similarity 28.6%; Pred. No. 14;
Matches 10; Conservative 5; Mismatches 3; Indels 17; Gaps 1;

QY  3 TSPWTTSS-----GLWPRLQKAA 20
DB  109 SSPWASCCSRSGLMVRRSQPASCSTWPLRLKLA 143

RESULT 4
US-08-856-444-2
; Sequence 2, Application US/08856444
; Patent No. 5959081
; GENERAL INFORMATION:
; APPLICANT: Lecka-Czernik, Beata
; TITLE OF INVENTION: No. 5959081el Zinc Binding LIM Protein S2-6
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benjamin Aaron Adler, Ph.D. J.D.
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/856,444
; FILING DATE: May 14, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5988
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 777-2321
; TELEFAX: (713) 777-6908
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: protein
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; IMMEDIATE SOURCE:
; POSITION IN GENOME:
; FEATURE:
; PUBLICATION INFORMATION:
; US-08-856-444-2

Query Match          31.3%; Score 48.5; DB 2; Length 343;
Best Local Similarity 52.6%; Pred. No. 36;
Matches 10; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

QY  1 LTSPWTTSSGLWPRLQKA 19
DB  230 LITAPWSTRQG---RLQQA 245

RESULT 5
US-08-554-612C-1
; Sequence 1, Application US/08554612C
; Patent No. 5747660
; GENERAL INFORMATION:
; APPLICANT: Orlicky, David
; TITLE OF INVENTION: PROTAGLANDIN F2' RECEPTOR REGULATORY
; TITLE OF INVENTION: PROTEIN AND THERAPEUTIC USES
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/554,612C
; FILING DATE: No. 5747660ember 6, 1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: 06519/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 322-5070
; TELEFAX: (415) 854-0875
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 879 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-554-612C-1

Query Match          31.3%; Score 48.5; DB 1; Length 879;
Best Local Similarity 37.9%; Pred. No. 1.1e+02;
Matches 11; Conservative 3; Mismatches 14; Indels 1; Gaps 1;
```

QY 2 LTSPWTTSSGLWPRLO-KAAEFKQLNOP 29
Db 248 VVSEWITEQGSQWQEIQEKAVEVATVVIQ 276

RESULT 6

US-09-583-110-5129
; Sequence 5129, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5129
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5129

Query Match 30.6%; Score 47.5; DB 4; Length 361;
Best Local Similarity 50.8%; Pred. No. 55;
Matches 11; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 WTTSSGLWPRLOKAAE-AFKQL 26
Db 26 WPTRPGSWPPOGKAAGAFQI 47

RESULT 7

US-09-252-991A-29866
; Sequence 29866, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29866
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29866

Query Match 30.6%; Score 47.5; DB 4; Length 611;
Best Local Similarity 44.4%; Pred. No. 1e+02;
Matches 12; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

QY 1 ILTSPWTTSSGLWPRLOKAAEFKQLN 27
Db 341 LLGSAGTQARGLWEFLRDAADS-KQLH 366

RESULT 8

US-09-881-710-12
; Sequence 12, Application US/09881710

; Patent No. 6673895
; GENERAL INFORMATION:
; APPLICANT: DESPRES, Phillipe
; APPLICANT: COURAGEOT, Marie-Pierre
; APPLICANT: DEUBEL, Vincent
; APPLICANT: CATTEAU, Adeline
; FILE REFERENCE: PRO-APOPTOTIC FRAGMENTS OF THE DENGUE VIRUS ENVELOPE GLYCOPROTEIN
; FILE REFERENCE: 209671USO
; CURRENT APPLICATION NUMBER: US/09/881,710
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/212,129
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Dengue virus
US-09-881-710-12

Query Match 30.3%; Score 47; DB 4; Length 39;
Best Local Similarity 43.8%; Pred. No. 4.6;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 TSPWTTSSGLWPRLOK 18
Db 16 TETWMSSEGAWKQIQK 31

RESULT 9

US-09-881-710-30
; Sequence 30, Application US/09881710
; Patent No. 6673895
; GENERAL INFORMATION:
; APPLICANT: DESPRES, Phillipe
; APPLICANT: COURAGEOT, Marie-Pierre
; APPLICANT: DEUBEL, Vincent
; APPLICANT: CATTEAU, Adeline
; TITLE OF INVENTION: PRO-APOPTOTIC FRAGMENTS OF THE DENGUE VIRUS ENVELOPE GLYCOPROTEIN
; FILE REFERENCE: 209671USO
; CURRENT APPLICATION NUMBER: US/09/881,710
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/212,129
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Dengue virus
US-09-881-710-30

Query Match 30.3%; Score 47; DB 4; Length 39;
Best Local Similarity 43.8%; Pred. No. 4.6;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 TSPWTTSSGLWPRLOK 18
Db 16 TETWMSSEGAWKQIQK 31

RESULT 10

US-09-881-710-28
; Sequence 28, Application US/09881710
; Patent No. 6673895
; GENERAL INFORMATION:
; APPLICANT: DESPRES, Phillipe
; APPLICANT: COURAGEOT, Marie-Pierre
; APPLICANT: DEUBEL, Vincent
; APPLICANT: CATTEAU, Adeline
; TITLE OF INVENTION: PRO-APOPTOTIC FRAGMENTS OF THE DENGUE VIRUS ENVELOPE GLYCOPROTEIN
; FILE REFERENCE: 209671USO
; CURRENT APPLICATION NUMBER: US/09/881,710

```
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/212,129
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Dengue virus
US-09-881-710-28

Query Match 30.3%; Score 47; DB 4; Length 40;
Best Local Similarity 43.8%; Pred. No. 4.7;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 TSPWTTSSGLWPRLQK 18
DB 16 TETWMSSEGAWKQIQK 31

RESULT 11
US-09-881-710-1
; Sequence 1, Application US/09881710
; Patent No. 6673895
; GENERAL INFORMATION:
; APPLICANT: DESPRES, Philippe
; APPLICANT: COURAGEOT, Marie-Pierre
; APPLICANT: DEUBEL, Vincent
; APPLICANT: CATTEAU, Adeline
; TITLE OF INVENTION: PRO-APOPTOTIC FRAGMENTS OF THE DENGUE VIRUS ENVELOPE GLYCOPROTEIN
; CURRENT APPLICATION NUMBER: US/09/881,710
; FILE REFERENCE: 209671US0
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/212,129
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Dengue virus
US-09-881-710-1

Query Match 30.3%; Score 47; DB 4; Length 76;
Best Local Similarity 43.8%; Pred. No. 10;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 TSPWTTSSGLWPRLQK 18
DB 52 TETWMSSEGAWKQIQK 67

RESULT 12
US-09-513-999C-7233
; Sequence 7233, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 7233
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 68
; OTHER INFORMATION: Xaa=Asp or Glu
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 69
; OTHER INFORMATION: Xaa=Ala or Asp or His or Asn or Pro or Ser or Thr or Tyr
US-09-513-999C-7233

Query Match 30.3%; Score 47; DB 4; Length 84;
Best Local Similarity 35.3%; Pred. No. 11;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 13 WPRLQKAAEAFKQLNQP 29
DB 30 WPRINSATGYQRMKP 46

RESULT 13
US-09-252-991A-20461
; Sequence 20461, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20461
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20461

Query Match 30.3%; Score 47; DB 4; Length 146;
Best Local Similarity 37.9%; Pred. No. 22;
Matches 11; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 ILTSPWTTSSGLWPRLQKAAEAFKQLNQP 29
DB 99 IACSPRASSAALAPRISRAFLWLSLNL 127

RESULT 14
US-09-187-999-2
; Sequence 2, Application US/09187999A
; Patent No. 6482646
; GENERAL INFORMATION:
; APPLICANT: Gindullis, Frank
; APPLICANT: Meier, Iris
; TITLE OF INVENTION: Plant Proteins that Interact with Nuclear Matrix
; TITLE OF INVENTION: Proteins and Function as Transcriptional Activators
; FILE REFERENCE: CL-1321
; CURRENT APPLICATION NUMBER: US/09/187,999A
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-187-999-2

Query Match 30.3%; Score 47; DB 4; Length 152;
Best Local Similarity 44.4%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
```

QY 5 PWTSSGLWPRLOKAAEA 22
 Db 21 PMTISFSINWPTQETRDA 38

RESULT 15
 US-09-252-991A-25179
 ; Sequence 25179, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 25179
 ; LENGTH: 210
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-25179

Query Match 30.3%; Score 47; DB 4; Length 210;
 Best Local Similarity 39.3%; Pred. No. 34;
 Matches 11; Conservative 2; Mismatches 7; Indels 8; Gaps 1;

Qy 3 TSPWTTSSGLWPR-----LQKAAEA 22
 Db 16 SGPWTRRSGWIRAGCAVTPLOPSCA 43

Search completed: November 10, 2004, 13:44:07
 Job time : 10.0173 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 31.1537 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-36

Perfect score: 155

Sequence: 1 ILTSPWTTSSGLWPRLQKAAEAFKQLNQ 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155	100.0	29	14	US-10-092-750-36
2	67	43.2	535	15	US-10-296-115-1110
3	55	35.5	259	16	US-10-767-701-40845
4	55	35.5	600	16	US-10-437-963-193474
5	53	34.2	460	16	US-10-437-963-192472
6	52	33.5	123	17	US-10-425-115-233344
7	52	33.5	531	14	US-10-369-493-8143
8	51	32.9	175	17	US-10-425-115-241808
9	51	32.9	793	14	US-10-104-047-2549
10	51	32.9	2131	16	US-10-408-765A-1179
11	50	32.3	108	17	US-10-425-115-301225
12	50	32.3	264	15	US-10-108-260A-3910
13	50	32.3	566	10	US-09-775-803-12

14	50	32.3	567	15	US-10-080-334-228	Sequence 228, Appl
15	50	32.3	567	15	US-10-037-417-105	Sequence 105, Appl
16	50	32.3	575	16	US-10-437-963-171554	Sequence 171554,
17	50	32.3	761	16	US-10-781-599-63	Sequence 63, Appl
18	49.5	31.9	175	17	US-10-425-115-34373	Sequence 34373, A
19	49.5	31.9	236	15	US-10-425-114-45426	Sequence 45426, A
20	49.5	31.9	240	17	US-10-425-115-257631	Sequence 257631,
21	49	31.6	136	15	US-10-424-599-167190	Sequence 167190,
22	49	31.6	268	14	US-10-107-096-2	Sequence 2,
23	49	31.6	331	14	US-10-156-761-10380	Sequence 10380, A
24	49	31.6	373	15	US-10-382-122A-43721	Sequence 43721, A
25	48.5	31.3	260	15	US-10-264-049-2527	Sequence 2527, Ap
26	48.5	31.3	320	16	US-10-437-963-119917	Sequence 119917,
27	48.5	31.3	343	15	US-10-396-260A-2	Sequence 2, Appl
28	48	31.0	106	15	US-10-424-599-168666	Sequence 168666,
29	48	31.0	186	15	US-10-424-599-217017	Sequence 217017,
30	47.5	30.6	220	16	US-10-437-963-192804	Sequence 192804,
31	47	30.3	220	16	US-10-311-213-18	Sequence 18, Appl
32	47	30.3	35	16	US-10-311-213-13	Sequence 13, Appl
33	47	30.3	39	9	US-09-881-710-12	Sequence 12, Appl
34	47	30.3	39	9	US-09-881-710-10	Sequence 30, Appl
35	47	30.3	39	15	US-10-634-895-12	Sequence 12, Appl
36	47	30.3	39	15	US-10-634-895-30	Sequence 30, Appl
37	47	30.3	39	16	US-10-311-213-16	Sequence 16, Appl
38	47	30.3	40	9	US-09-881-710-28	Sequence 28, Appl
39	47	30.3	40	15	US-10-634-895-28	Sequence 28, Appl
40	47	30.3	40	16	US-10-311-213-1	Sequence 1, Appl
41	47	30.3	40	16	US-10-311-213-20	Sequence 20, Appl
42	47	30.3	47	16	US-10-311-213-9	Sequence 9, Appl
43	47	30.3	76	9	US-09-881-710-1	Sequence 1, Appl
44	47	30.3	76	15	US-10-634-895-1	Sequence 1, Appl
45	47	30.3	78	17	US-10-425-115-189993	Sequence 189993,

ALIGNMENTS

RESULT 1
US-10-092-750-36
; Sequence 36, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-XL
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-36

Query Match 100.0%; Score 155; DB 14; Length 29;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILTSPWTTSSGLWPRLQKAAEAFKQLNQ 29

Db 1 ILTSPWTTSSGLWPRLQKAAEAFKQLNQ 29

RESULT 2

US-10-296-115-1110
; Sequence 1110, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:

```
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296.115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1110
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-115-1110

Query Match      43.2%; Score 67; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 RLQAAEAFKQLNQ 38
DB 92 RLQAAEAFKQLNQ 105

RESULT 3
US-10-767-701-40845
; Sequence 40845, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(533535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 40845
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C10916_1.pep
US-10-767-701-40845

Query Match      35.5%; Score 55; DB 16; Length 259;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 SPWTTSSGLWPRLOKA 19
DB 37 SPGATATSLWPRLOKA 52

RESULT 4
US-10-437-963-193474
; Sequence 193474, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 193474
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_89609C.1.pep
US-10-437-963-193474

Query Match      34.2%; Score 53; DB 16; Length 460;
Best Local Similarity 43.5%; Pred. No. 74;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 6 WTTSSGLWPRLOKAAEAFKQLNQ 28
DB 408 WTASALLWPRLLAALESDESLGR 430

RESULT 6
US-10-425-115-233344
; Sequence 233344, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 233344
; LENGTH: 123
; TYPE: PRT
```

```
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 193474
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_89609C.1.pep
US-10-437-963-193474

Query Match      35.5%; Score 55; DB 16; Length 600;
Best Local Similarity 58.8%; Pred. No. 52;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 6 WTTSSGLWPRLOKAAEAFKQLNQ 22
DB 429 WTVSARLWPRLLADAIEA 445

RESULT 5
US-10-437-963-192472
; Sequence 192472, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 192472
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88698C.1.pep
US-10-437-963-192472

Query Match      34.2%; Score 53; DB 16; Length 460;
Best Local Similarity 43.5%; Pred. No. 74;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 6 WTTSSGLWPRLOKAAEAFKQLNQ 28
DB 408 WTASALLWPRLLAALESDESLGR 430

RESULT 6
US-10-425-115-233344
; Sequence 233344, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 233344
; LENGTH: 123
; TYPE: PRT
```

```
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_144402C.1.pap
US-10-425-115-233344

Query Match      33.5%; Score 52; DB 17; Length 123;
Best Local Similarity 55.6%; Pred. No. 26;
Matches 10; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

QY 5 PWTSSGL--WPRLOKAA 20
   |||||:|||||
Db 52 PWTPKNGLGWPRFPKKA 69
   |||||:|||||

RESULT 7
US-10-369-493-8143
; Sequence 8143, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8143
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Thermobifida fusca
US-10-369-493-8143

Query Match      33.5%; Score 52; DB 14; Length 531;
Best Local Similarity 45.5%; Pred. No. 1.2e+02;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 5 PWTSSGLWPRLOKAAAFKQL 26
   |||||:|||||
Db 379 PWTSLGYGMPEKTAEWRNL 400
   |||||:|||||

RESULT 8
US-10-425-115-241808
; Sequence 241808, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 241808
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: {1}..(175)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MMT4577_152107C.1.pap
US-10-425-115-241808
```

```
Query Match      32.9%; Score 51; DB 17; Length 175;
Best Local Similarity 29.3%; Pred. No. 52;
Matches 12; Conservative 4; Mismatches 18; Indels 1;

QY 7 TTSSGLWPRL-----OKAAEAFKQLNOP 29
   |||||:|||||
Db 129 TLPAGLWPEIYRLQWKEEYMGATLVDTQKQADALRTLRQ 169
   |||||:|||||

RESULT 9
US-10-104-047-2549
; Sequence 2549, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR FILING DATE: 2002-03-25
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2549
; LENGTH: 793
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2549

Query Match      32.9%; Score 51; DB 14; Length 793;
Best Local Similarity 36.0%; Pred. No. 2.5e+02;
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 4 SPWTTSSGLWPRLOKAAEAFKQLNQ 28
   |||||:|||||
Db 709 SPWPVSSALTARLRRLVTYQRCNR 733
   |||||:|||||

RESULT 10
US-10-408-765A-1179
; Sequence 1179, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1179
; LENGTH: 2131
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1179

Query Match      32.9%; Score 51; DB 16; Length 2131;
Best Local Similarity 36.0%; Pred. No. 6.8e+02;
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 4 SPWTTSSGLWPRLOKAAEAFKQLNQ 28
   |||||:|||||
Db 1102 SPWPVSSALTARLRRLVTYQRCNR 1126
   |||||:|||||
```

```
RESULT 11
US-10-425-115-301225
; Sequence 301225, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 301225
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(108)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MET4577_37793C.1.pcp
US-10-425-115-301225

Query Match      32.3%; Score 50; DB 17; Length 108;
Best Local Similarity 52.9%; Pred. No. 43;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      3 TSPWTTSSGLWPRLQKA 19
Db      61 TPTPTTSSRMWMPQSA 77

RESULT 12
US-10-108-260A-3910
; Sequence 3910, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3910
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3910

Query Match      32.3%; Score 50; DB 15; Length 264;
Best Local Similarity 37.0%; Pred. No. 1.le+02;
Matches 10; Conservative 6; Mismatches 9; Indels 2; Gaps 1;

QY      5 PWTTSGLWPRLQ--KAAEAFKQLNQP 29
Db      204 PWTAAAWGELLEGAKASAKGEQVRDP 230

RESULT 13
US-09-775-803-12
; Sequence 12, Application US/09775803
; Publication No. US20030167487A1
; GENERAL INFORMATION:
; APPLICANT: COR Therapeutics, Inc.
; APPLICANT: Ramakrishnan, Vanitha
; APPLICANT: Phillips, David
; TITLE OF INVENTION: Transgenic Animals Having a Modified Glycoprotein V
; TITLE OF INVENTION: Gene and Methods for Their Use
```

```
; FILE REFERENCE: 44481-5044-US
; CURRENT APPLICATION NUMBER: US/09/775,803
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/17594
; PRIOR FILING DATE: 1999-08-04
; PRIOR APPLICATION NUMBER: US 60/109,797
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-775-803-12

Query Match      32.3%; Score 50; DB 10; Length 566;
Best Local Similarity 57.1%; Pred. No. 2.4e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      4 SPWTTSSGLWPRLQ 17
Db      421 NPWLCDGGLWPFLQ 434

RESULT 14
US-10-080-334-228
; Sequence 228, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zerhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; TITLE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
```

```
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 228
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-080-334-228

Query Match      32.3%; Score 50; DB 15; Length 567;
Best Local Similarity 57.1%; Pred. No. 2.4e+02;
Matches      8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      4 SPWTTSSGLWPRLQ 17
Db      421 NPWLCDGGLWPFLLQ 434

RESULT 15
US-10-037-417-105
; Sequence 105, Application US/10037417
; Publication No. US20040052806A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Rothenberg, Mark
; APPLICANT: Store, David J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Anderson, David W
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Miller, Charles E
; APPLICANT: Eisen, Andrew J
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-235
; CURRENT APPLICATION NUMBER: US/10/037,417
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/260,018
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: 60/260,360
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/272,411
; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/272,817
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/291,186
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/303,231
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/305,060
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/318,405
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/318,700
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 227
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 105
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-417-105

Query Match      32.3%; Score 50; DB 15; Length 567;
Best Local Similarity 57.1%; Pred. No. 2.4e+02;
Matches      8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      4 SPWTTSSGLWPRLQ 17
Db      421 NPWLCDGGLWPFLLQ 434

Search completed: November 11, 2004, 01:28:18
Job time : 32.2037 secs
```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 6.46114 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-36

Perfect score: 155
Sequence: 1 ILTSPWTTSSGLWPRLQKAAEAFKQLNQP 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*

```
1:  _pir1:*
```

2: pir2: *

```
3: dir3:*
```

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	50	32.3	282	2	T36983	probable transposase	
2	50	32.3	401	2	A12255	two-component sens	
3	50	32.3	480	1	IBEG	hydroxymethylbilan	
4	49	31.6	252	2	C85723	probable enzyme Z2	
5	49	31.6	232	2	P90894	probable enzyme [i	
6	49	31.6	222	2	B64906	biotin biosynthesi	
7	49	31.6	378	2	A10163	probable periplasm	
8	49	31.6	424	2	S52847	egg membrane prote	
9	49	31.6	698	2	H71535	hypothetical prote	
10	49	31.6	831	2	T49721	hypothetical prote	
11	48	31.0	454	2	F90602	aminopeptidase (le	
12	48	31.0	688	2	T37923	hypothetical prote	
13	47.5	30.6	361	2	D95106	conserved hypothet	
14	47.5	30.5	450	2	G81536	hypothetical prote	
15	47	30.3	233	2	H81890	conserved hypothet	
16	47	30.3	281	2	S41854	phosphoprotein pho	
17	47	30.3	441	2	T49265	hypothetical prote	
18	47	30.3	775	2	A47311	polyprotein(C, E,	
19	47	30.3	792	2	C32401	genome polyprotein	
20	47	30.3	792	2	B32401	genome polyprotein	
21	47	30.3	792	2	A32401	genome polyprotein	
22	47	30.3	821	2	S39983	epas protein - mou	
23	47	30.3	822	2	I38728	epidermal growth f	
24	47	30.3	1286	1	GNWVWP	genome polyprotein	
25	47	30.3	1785	2	T22595	hypothetical prote	
26	46.5	30.0	305	2	H82983	probable transcrip	
27	46.5	30.0	466	2	S42210	ISG-54K protein -	
28	46.5	30.0	599	2	T51548	hypothetical prote	
29	46	29.7	103	2	S15151	hypothetical prote	

	30	46	29.7	169	2	P84716
	31	46	29.7	176	2	T47216
	32	46	29.7	340	2	H91084
	33	46	29.7	365	2	A85930
	34	46	29.7	413	2	AG2363
	35	46	29.7	419	2	F72089
	36	46	29.7	419	2	F86534
	37	46	29.7	446	2	H65060
	38	46	29.7	691	2	D96805
	39	46	29.7	775	2	F98136
	40	46	29.7	800	2	AH3151
	41	46	29.7	867	2	AH0437
	42	46	29.7	963	2	A53984
	43	46	29.7	983	2	AG2381
	44	46	29.7	998	1	QOBEB1
	45	46	29.7	998	2	S41397

ALIGNMENTS

```

RESULT 1
T36983
probable transposase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Ju-1999
C:Accession: T36983
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.;
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21618
A:Accession: T36983
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-282 <OLI>
A:Cross-references: UNIPROT:Q9R192; EMBL:AL109949; PIDN:CAB52897.1; G
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB3:SCJ11.12
C:Superfamily: Streptomyces coelicolor probable transposase SC699.365

```

Query Match 32.3%; Score 50; DB 2; Length 282;
Best Local Similarity 53.3%; Pred. No. 16;
Matches 8: Conservative 2; Mismatches 5; Indels

```

Qy      3 TSPWTTSSGLWPLQ 17
         |||| | ||| ::
Db     16 TSPWIVSDELWDRVE 30

```

RESULT 2

A:Accession: AI2255
A:Gene: all3600
A:Superfamily: sensory transduction system regulatory protein homolog

A:Molecule type: DNA
A:Status: Preliminary

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
A:Reference numbers: AB1807; MUID:21595285; PMID:11759840

A:Cross-references: UNIPROT:O8VR50; GB:BA000019; PIDN:BAF5299.1; PID:gi7132733; GSPDB:1-401 <KUR>
A:Residues: 1-401

C:Genetics:
C:Experimental source: strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Species: Nostoc sp. PCC 7120
C:Accession: AI2255
E:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iritiguch, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yanada, M.; Yasuda, M.; Tabata, D.N.A. Res. 8, 205-213, 2001

E:two-component sensor histidine kinase all3600 [imported] - Nostoc sp. (strain PCC 7120)

Query Match 32.3%; Score 50; DB 2; Length 401;
Best Local Similarity 40.9%; Pred. No. 23;

Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 7 TTSSGLWPRLOKAAEAFKQLNQ 28
 DB 96 TQLXNLWPRWQAADTVAKLQE 117

RESULT 3
 IREG
 hydroxymethylbilane synthase (EC 4.3.1.8) precursor - Euglena gracilis
 A:Alternate names: porphobilinogen deaminase; pre-uroporphyrinogen synthase
 C:Species: Euglena gracilis
 C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
 C:Accession: S06109
 R:Sharif, A.L.; Smith, A.G.; Abell, C.
 Eur. J. Biochem. 184, 353-359, 1989
 A:Title: Isolation and characterisation of a cDNA clone for a chlorophyll synthesis enzyme isolated with a very long transit peptide in Euglena.
 A:Reference number: S06109; MUID:90005485; PMID:2477247
 A:Accession: S06109
 A:Molecule type: mRNA
 A:Residues: 1-480 <SHA>
 A:Cross-references: UNIPROT:P13446; GB:X15743; NID:g18411; PIDN:CAA33759.1; PID:g18412
 A:Note: Part of this sequence, including the amino end of the mature protein, was confirmed. 175-Ala was also found
 C:Function:
 A:Description: catalyzes the stepwise polymerization of four molecules of porphobilinogen
 A:Pathway: porphyrin biosynthesis
 A:Note: acting with uroporphyrinogen-III synthase (cosynthase), which cyclizes hydroxymethylbilane to uroporphyrinogen-III
 C:Superfamily: hydroxymethylbilane synthase
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase; chlorophyll biosynthesis; chloroplast; F11-139/Domain: transit peptide (chloroplast) #status predicted <TMP>
 F:140-480/Product: hydroxymethylbilane synthase #status experimental <MAT>
 F:395/Modified site: dipyrrolylmethanemethyl (Cys) (covalent) #status predicted

Query Match 32.3%; Score 50; DB 1; Length 480;
 Best Local Similarity 55.6%; Pred. No. 28;
 Matches 10; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 5 PWTSSGLWPRLOKAAEA 22
 DB 435 PW--SGRTWPOLQKESHA 450

RESULT 4
 C85723
 probable enzyme Z2186 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C:Accession: C85723
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Ziegler, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Pocamousis, K.; Apodaca, Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: C85723
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-252 <STO>
 A:Cross-references: UNIPROT:Q8XAZ2; GB:AE005174; NID:g12515147; PIDN:AAG56247.1; GSPDB:G12515147
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z2186
 C:Superfamily: bioC protein; bioC homology

Query Match 31.6%; Score 49; DB 2; Length 252;
 Best Local Similarity 47.6%; Pred. No. 19;
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 6 WTTSSGLWPRLOKAAEAFKQL 26
 DB 196 WVTATGLRPWLQDLTSEQQ 216

RESULT 7
 A10163
 probable periplasmic binding protein YP01343 [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
 C:Accession: A10163
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, F. Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A10163

RESULT 5
 F90894
 probable enzyme [imported] - Escherichia coli (strain O157:H7, substrain RMD 050952)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
 C:Accession: F90894
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome reference number: A96623; MUID:21156231; PMID:11258796
 A:Accession: F90894
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-252 <HAY>
 A:Cross-references: UNIPROT:Q8XAZ2; GB:BA000007; PIDN:BA035549.1; PID:g13361592; GSPDB:G13361592
 A:Experimental source: strain O157:H7, substrain RMD 050952
 C:Genetics:
 A:Gene: BC2126
 C:Superfamily: bioC protein; bioC homology

Query Match 31.6%; Score 49; DB 2; Length 252;
 Best Local Similarity 47.6%; Pred. No. 19;
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 6 WTTSSGLWPRLOKAAEAFKQL 26
 DB 196 WVTATGLRPWLQDLTSEQQ 216

RESULT 6
 B64906
 biotin biosynthesis protein homolog b1519 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
 C:Accession: B64906
 R:Plattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col, A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: B64906
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <BLAT>
 A:Cross-references: UNIPROT:P76145; GB:AE000249; GB:U00096; NID:g1787790; PIDN:AAC74592.1
 A:Experimental source: strain K-12, substrain MG1655
 C:Superfamily: bioC protein; bioC homology
 F:32-130/Domain: bioC homology <BIOC>

Query Match 31.6%; Score 49; DB 2; Length 252;
 Best Local Similarity 47.6%; Pred. No. 19;
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 6 WTTSSGLWPRLOKAAEAFKQL 26
 DB 196 WVTATGLRPWLQDLTSEQQ 216

RESULT 7
 A10163
 probable periplasmic binding protein YP01343 [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
 C:Accession: A10163
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, F. Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A10163

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-378 <UR>
A;Cross-references: UNIPROT:Q8ZGF8; GB:AL590842; PIDN:CAC90172.1; PID:g15979391; GSPDB:G
C;Genetics:
A;Gene: YP01343

Query Match      31.6%; Score 49; DB 2; Length 378;
Best Local Similarity 39.1%; Pred. No. 30;
Matches 9; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 4 SPWTTSSGLWPRLOKAAEAKQL 26
Db 84 SDWSLRTSQWPRYQAALRLSQL 106

RESULT 8
S52847
egg membrane protein ZP3 (clone pCOV638) - common carp (fragment)
C;Species: Cyprinus carpio (common carp)
C;Date: 09-Jun-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C;Accession: S52847
R;Chang, Y.; Wang, S.; Tsao, C.; Huang, F.
submitted to the EMBL Data Library, April 1995
A;Description: Structural analysis and expression of carp ZP3 gene.
A;Reference number: S52847
A;Accession: S52847
A;Molecule type: mRNA
A;Residues: 1-424 <CHA>
A;Cross-references: UNIPROT:Q92027; EMBL:Z48972; NID:g763077; PIDN:CAA88836.1; PID:g7630
C;Superfamily: sperm-binding glycoprotein ZP3; ZP domain homology
P;110-364/Domain: ZP domain homology <ZPH>

Query Match      31.6%; Score 49; DB 2; Length 424;
Best Local Similarity 45.0%; Pred. No. 34;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 ILTSPWTTSSGLWPRLOKAA 20
Db 38 VIRVPAVSSEGLWKTLOKAS 57

RESULT 9
H71535
hypothetical protein CT283 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
C;Accession: H71535
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trach
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: H71535
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-698 <ARN>
A;Cross-references: UNIPROT:O84285; GB:AB001300; GB:AB001273; NID:G3328682; PIDN:AAC6787
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: CT283
C;Superfamily: Chlamydia trachomatis hypothetical protein CT283

Query Match      31.6%; Score 49; DB 2; Length 698;
Best Local Similarity 33.3%; Pred. No. 60;
Matches 10; Conservative 8; Mismatches 8; Indels 4; Gaps 1;

QY 4 SPWTTSSGLW----PRLOKAAEAKQLNQ 29
Db 135 APFISSEVWKSAPQLEAFHIFQQTDP 164

RESULT 10
T49721
hypothetical protein B23L21.340 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T49721
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49721
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-831 <SCH>
A;Cross-references: UNIPROT:Q9P5J5; EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.340
A;Experimental source: BAC clone B23L21; strain OR74A
C;Genetics:
A;Gene: NCSP:B23L21.340
A;Map position: 6
A;Introns: 31/3; 629/3

Query Match      31.6%; Score 49; DB 2; Length 831;
Best Local Similarity 38.1%; Pred. No. 73;
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 ILTSPWTTSSGLWPRLOKAAE 21
Db 770 LITSPSPAELWPKFEEMAK 790

RESULT 11
F90602
aminopeptidase (leucine aminopeptidase) (lap) [imported] - Mycoplasma pulmonis (strain U
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: F90602
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: F90602
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-454 <KUR>
A;Cross-references: UNIPROT:Q98PJ6; GB:AL445566; PID:g14090141; PIDN:CAC13899.1; GSPDB:G
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPV 7260
A;Genetic code: SGC3
C;Superfamily: Cytosol aminopeptidase

Query Match      31.0%; Score 48; DB 2; Length 454;
Best Local Similarity 36.8%; Pred. No. 52;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 3 TSPWTTSSGLWPRLOKAAE 21
Db 354 TGVWSINDSKWEKLEKASK 372

RESULT 12
T37923
hypothetical protein SPAC18G6.10 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T37923
R;Connor, R.; Churcher, C.M.; Barrrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
submitted to the EMBL Data Library, December 1999
A;Reference number: Z21754
A;Accession: T37923
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-688 <CON>
A;Cross-references: UNIPROT:Q10109; EMBL:Z68198; PIDN:CAA92388.1; GSPDB:GN000066; SPDB:SP
A;Experimental source: strain 972h-; cosmid c18G6
C;Genetics:
```

A:Gene: SPDB:SPAC18G6.10
A:Map position: 1

Query Match 31.0%; Score 48; DB 2; Length 688;
Best Local Similarity 50.0%; Pred. No. 82;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 ILTSPWTTSSGLWP 14

DB 390 VLVEPWTSLSLGF 403

RESULT 13

D95106
conserved hypothetical protein SP0921 [imported] - Streptococcus pneumoniae (strain TIGR)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004

C:Accession: D95106

R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle,

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Lofthus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: D95106

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-361 <KUR>

A:Cross-references: UNIPROT:Q97RA4; GB:AE005672; PIDN:AAK75045.1; PID:gl4972395; GSPDB:G

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0921

C:Superfamily: uncharacterized conserved protein SP0921

Query Match 30.6%; Score 47.5; DB 2; Length 361;

Best Local Similarity 50.0%; Pred. No. 47;

Matches 11; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 WTTSSGLWPRLQKAAE-AFKQL 26

DB 26 WPTFGSNWPFQGAKEAFTQI 47

RESULT 14

G83536
hypothetical protein PA0881 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C:Accession: G83536

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: G83536

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-450 <STO>

A:Cross-references: UNIPROT:Q91564; GB:AE004522; GB:AE004091; NID:g9946768; PIDN:AAQ0427

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA0881

Query Match 30.6%; Score 47.5; DB 2; Length 450;

Best Local Similarity 44.4%; Pred. No. 60;

Matches 12; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

QY 1 ILTSPWTTSSGLWPRLQKAAE-AFKQLN 27

DB 180 LLGSAGTGAAGLWFLRDAADS-KQLH 205

RESULT 15

H81690

conserved hypothetical protein frameshift TC0556 [imported] - Chlamydia muridarum (strain

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004

C:Accession: H81690

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heigelberg, J.F.; White, O.; Hickey, E

.C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: H81690

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-253 <TET>

A:Cross-references: UNIPROT:Q9PKB1; GB:AE002323; GB:AE002160; NID:g7190585; PIDN:AAF3939;

A:Experimental source: strain Nigg (MoPn)

C:Genetics:

A:Gene: TC0556

Query Match 30.3%; Score 47; DB 2; Length 253;

Best Local Similarity 33.3%; Pred. No. 37;

Matches 10; Conservative 7; Mismatches 9; Indels 4; Gaps 1;

QY 4 SPWTTSSGLW----PRLOKAAEAFKQLNQP 29

DB 127 APFISSEEVWKKSSAPQLRDAFHIFQLENP 156

Search completed: November 10, 2004, 13:40:44

Job time : 7.46114 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 35.4111 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-36
Perfect score: 155
Sequence: 1 ILTSPWTTSSGLWPLRQKAAEFKQLNP 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	43.2	400	Q91V77	Q91V77 mus musculus
2	67	43.2	715	Q9H6K9	Q9H6K9 homo sapien
3	67	43.2	729	Q9H6K3	Q9H6K3 mus musculus
4	67	43.2	743	Q9H6S3	Q9H6S3 homo sapien
5	56	36.1	799	Q72BM7	Q72BM7 desulfobiv
6	56	36.1	799	RAS96086	RAS96086 desulfovi
7	55	35.5	1557	Q6ZUB3	Q6ZUB3 homo sapien
8	55	35.5	1557	BAC8313	BAC8313 homo sapi
9	54	34.8	707	Q7R724	Q7R724 plasmodium
10	53	34.2	251	Q81W90	Q81W90 bacillus an
11	53	34.2	251	RAT33212	RAT33212 bacillus
12	53	34.2	460	Q9LDD9	Q9LDD9 oryza sativ
13	52.5	33.9	280	Q7U758	Q7U758 synchococc
14	52.5	33.9	413	Q8SV18	Q8SV18 enccephalito
15	51.5	33.2	306	Q88B18	Q88B18 pseudomonas
16	51.5	33.2	308	Q88CC0	Q88CC0 pseudomonas
17	51.5	33.2	598	Q993H6	Q993H6 callitrichi
18	51.5	33.2	1503	Q6SLB9	Q6SLB9 gibberella
19	51.5	33.2	1503	RAR30119	RAR30119 gibberell
20	51	32.9	347	Q05160	Q05160 synchococc
21	51	32.9	2713	CHD6 HUMAN	CHD6 homo sapien
22	50.5	32.6	328	Q7VSC4	Q7VSC4 bordetella
23	50.5	32.6	328	Q7WKE7	Q7WKE7 bordetella
24	50	32.3	264	Q8N816	Q8N816 homo sapien
25	50	32.3	282	Q9RI92	Q9RI92 streptomyces
26	50	32.3	401	Q8YR50	Q8YR50 arabidopsis
27	50	32.3	475	Q8YV88	Q8YV88 ralestonia s
28	50	32.3	480	HEM3 EUGGR	HEM3 euglena gra
29	50	32.3	506	Q8LJA3	Q8LJA3 oryza sativ
30	50	32.3	567	GPV RAT	GPV rattus norv
31	50	32.3	567	Q9QZU3	Q9QZU3 mus musculus

32	50	32.3	761	2	Q9WT88	Q9WT88 tt virus. o
33	50	32.3	3163	2	Q7SAD6	Q7SAD6 neurospora
34	49.5	31.9	367	2	Q6IX33	Q6IX33 hypholoma f
35	49.5	31.9	367	2	AAC37495	AAC37495 hypholoma
36	49.5	31.9	510	2	Q27072	Q27072 taenia soli
37	49.5	31.9	595	2	Q7S925	Q7S925 neurospora
38	49	31.6	155	2	Q6XI15	Q6XI15 drosophila
39	49	31.6	155	2	AAR10040	AAR10040 drosophila
40	49	31.6	186	2	Q6XJ13	Q6XJ13 drosophila
41	49	31.6	186	2	Q9W3W8	Q9W3W8 drosophila
42	49	31.6	186	2	AAR09689	AAR09689 drosophila
43	49	31.6	192	2	Q7PDX3	Q7PDX3 anopheles g
44	49	31.6	251	1	TAM_ECO57	Q8XAZ2 escherichia
45	49	31.6	251	1	TAM_ECOLI	P76145 escherichia

ALIGNMENTS

RESULT 1				
ID	Q91V77	PRELIMINARY;	PRT;	400 AA.
AC	Q91V77;			
DT	01-DEC-2001 (TRENBLrel. 19, Created)			
DT	01-DEC-2001 (TRENBLrel. 19, Last sequence update)			
DT	01-MAR-2004 (TRENBLrel. 26, Last annotation update)			
DE	Eps812 protein.			
GN	Name=Eps812;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=wik FVB/N;			
RC	TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diachenko L., Marudina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
RA	Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E.,			
RA	Jones S.J., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			

Query Match 43.2%; Score 67; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RX MEDLINE=2388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Souffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RC Strausberg R.;
RL Submitted (MAR-2001) to the ENBL/GenBank/DBJ databases.
DR EMBL; BC005492; AAH05492.1; -.
DR HSSP; Q08509; 110C.
DR MGD; MGI:2138828; Eps812.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR011036; PH related.
DR InterPro; IPR006020; PTB_P1D.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3_1; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS50002; SH3; 1.
DR Receptor; SH3 domain.
KW Receptor.
SQ SEQUENCE 729 AA; 82229 MW; A11DE82FF2C0BD18 CRC64;

Query Match 43.2%; Score 67; DB 2; Length 729;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 RLQKAAEAFKQLNQ 28
Db 275 RLQKAAEAFKQLNQ 288

RESULT 4
Q9H6S3 PRELIMINARY; PRT; 743 AA.
AC Q9H6S3
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ21935.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RL Submitted (AUG-2000) to the ENBL/GenBank/DBJ databases.
DR EMBL; AK025588; BAB15180.1; -.
DR HSSP; Q08509; 110C.
DR Genew; HGNC:21296; EFS812.

QY 15 RLQKAAEAFKQLNQ 28
Db 272 RLQKAAEAFKQLNQ 285

RESULT 3
Q99K30 PRELIMINARY; PRT; 729 AA.
AC Q99K30
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Epidermal growth factor receptor pathway substrate 8-like protein
DE 2.
GN Name=Eps812;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mix FVB/N;

```

```

DR InterPro; IPR011036; PH related.
DR InterPro; IPR006020; PTE_PID.
DR InterPro; IPR001452; SH3.
DR Pfam; PF000018; SH3_1; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS00002; SH3; 1.
DR SH3 domain.
KW SH3 domain.
SQ SEQUENCE 743 AA; 83792 MW; DB01D8F6363A1F80 CRC64;

Query Match 43.2%; Score 67; DB 2; Length 743;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 RLOKAAEAFKQLNQ 28
Db 300 RLOKAAEAFKQLNQ 313

RESULT 5
Q72BM7 Q72BM7 PRELIMINARY; PRT; 799 AA.
AC Q72BM7
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE DNA ligase, NAD-dependent (EC 6.5.1.2).
GN NameLig; OrderedLocusNames=DVU1608;
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15077118; DOI=10.1038/nbr959;
RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N., Methe B.A., Brinkac L.M.,
RA Donnelly S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Davidson T.M., Zafar N., Zhou L., Radune D.,
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RA "The genome sequence of the anaerobic, sulfate-reducing bacterium
RT Desulfovibrio vulgaris Hildenborough."
RL Nat. Biotechnol. 22:554-559(2004).
CC -1- FUNCTION: This protein catalyzes the formation of phosphodiester
CC linkages between 5'-phosphoryl and 3'-hydroxyl groups in double-
CC stranded DNA using NAD as a coenzyme and as the energy source for
CC the reaction. It is essential for DNA replication and repair of
CC damaged DNA (By similarity).
CC -1- CATALYTIC ACTIVITY: NAD(+) + {deoxyribonucleotide}(N) +
CC {deoxyribonucleotide}(M) = AMP + nicotinamide nucleotide +
CC {deoxyribonucleotide}(N+M).
CC -1- SIMILARITY: Belongs to the NAD-dependent DNA ligase family.
DR EMBL; AE017314; AAS96086.1; -.
DR TIGR; DVU1608; -.
DR GO; GO:0003911; F:DNA ligase (NAD+) activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR001679; DNALigase.
DR InterPro; IPR004150; DNA_ligase_OB.
DR InterPro; IPR000445; HhH.
DR InterPro; IPR003583; HhH_1.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR InterPro; IPR010994; Ruva_2_like.
DR InterPro; IPR004149; Znf_DNALigase_C4.
DR Pfam; PF00533; BRCT; 1.
DR Pfam; PF01653; DNA_ligase_aden; 1.
DR Pfam; PF03120; DNA_ligase_OB; 1.
DR Pfam; PF03119; DNA_ligase_ZBD; 1.
DR Pfam; PF00633; HhH; 1.

ProDom; PD003944; DNALigase; 1.
DR SMART; SM00232; BRCT; 1.
DR SMART; SM00278; HhH; 3.
DR TIGR; TIGR00532; LIGANC; 1.
DR TIGR; TIGR00575; hnlj; 1.
DR PROSITE; PS01056; DNALIGASE_N2; 1.
DR PROSITE; PS01172; BRCT; 1.
DR PROSITE; PS01056; DNALIGASE_N2; 1.
KW Complete proteome; DNA repair; DNA replication; Ligase; NAD.
SQ SEQUENCE 799 AA; 86098 MW; 535A6009B6B60EA3 CRC64;

Query Match 36.1%; Score 56; DB 2; Length 799;
Best Local Similarity 62.5%; Pred. No. 48;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 TSPWTTSSGLWPRLQK 18
Db 270 TAPWTHSGLMARLRE 285

RESULT 6
AAS96086 PRELIMINARY; PRT; 799 AA.
ID AAS96086
AC AAS96086;
DT 27-APR-2004 (TrEMBLrel. 27, Created)
DT 27-APR-2004 (TrEMBLrel. 27, Last sequence update)
DT 11-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE DNA ligase, NAD-dependent (EC 6.5.1.2).
GN LIGA OR DVU1608.
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15077118;
RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N., Methe B.A., Brinkac L.M.,
RA Nelson W.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Peterson J.D., Davidson T.M., Zafar N., Zhou L., Radune D.,
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RA "The genome sequence of the anaerobic, sulfate-reducing bacterium
RT Desulfovibrio vulgaris Hildenborough."
RL Nat. Biotechnol. 22:554-559(2004).
DR EMBL; AE017314; AAS96086.1; -.
DR TIGR; DVU1608; -.
DR LIGase.
KW Ligase.
SQ SEQUENCE 799 AA; 86098 MW; 535A6009B6B60EA3 CRC64;

Query Match 36.1%; Score 56; DB 2; Length 799;
Best Local Similarity 62.5%; Pred. No. 48;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 TSPWTTSSGLWPRLQK 18
Db 270 TAPWTHSGLMARLRE 285

RESULT 7
Q6ZUB3 PRELIMINARY; PRT; 1557 AA.
ID Q6ZUB3
AC Q6ZUB3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ43854.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Iehli S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kaneshiro K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK125842; BAC6313.1; -.
DR InterPro: IP0005112; dENN.
DR InterPro: IP0001194; dENN.
DR InterPro: IP0002885; PPR.
DR Pfam: PF02141; dENN; 1.
DR Pfam: PF01535; PPR; 1.
DR PROSITE: PS00947; dENN; 1.
DR PROSITE: PS02111; dENN; 1.
SQ SEQUENCE 1557 AA; 173472 MW; 986E2B00DDEBDC11 CRC64;

Query Match 35.5%; Score 55; DB 2; Length 1557;
Best Local Similarity 36.4%; Pred. No. 1.3e+02;
Matches 12; Conservative 5; Mismatches 12; Indels 4; Gaps 2;

QY 1 ILTSPW--TTSSG--LWPRLOKAAEFKQLNOP 29
Db 513 VLESFWSSTRSGIFLTKVRNVVRGLAQRQP 545

RESULT 8
BAC6313 PRELIMINARY; PRT; 1557 AA.
ID BAC6313;
AC BAC6313;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE CDNA FLJ43854 fis, clone TEST14007064.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primata; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Iehli S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kaneshiro K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RL "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK125842; BAC6313.1; -.
SQ SEQUENCE 1557 AA; 173472 MW; 986E2B00DDEBDC11 CRC64;

Query Match 35.5%; Score 55; DB 2; Length 1557;
Best Local Similarity 36.4%; Pred. No. 1.3e+02;
Matches 12; Conservative 5; Mismatches 12; Indels 4; Gaps 2;

QY 1 ILTSPW--TTSSG--LWPRLOKAAEFKQLNOP 29
Db 513 VLESFWSSTRSGIFLTKVRNVVRGLAQRQP 545

RESULT 9
Q7R724 PRELIMINARY; PRT; 707 AA.
ID Q7R724;
AC Q7R724;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Hypothetical protein (Fragment).
GN Name=PY07765;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN NCBI_TaxID=73239;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=1236885;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feidbyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519 (2002).
CC -!- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
CC EMBL: AABL01002913; EAA20293.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 707
SQ SEQUENCE 707 AA; 77054 MW; DC40A56BA854AEFB CRC64;

Query Match 34.8%; Score 54; DB 2; Length 707;
Best Local Similarity 50.0%; Pred. No. 82;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 ILTSPWTTSSGLWPRLOKAA 20
Db 304 VLAWPTLPPSWRRRLQAA 323

RESULT 10
Q81W90 PRELIMINARY; PRT; 251 AA.
ID Q81W90; QGHU4; Q6KNM1;
AC Q81W90; QGHU4; Q6KNM1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Prphage Lambda02. Ctp protease family protein
GN OrderedAccession=BA4092, BAS3804; ORFNames=GBAA4092;
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate Porton;
RX MEDLINE=22608414; PubMed=1271629; DOI=10.1038/nature01586;
RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Ostad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beaman M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,
RA Salzberg S.L., Thomson B., Friedlander A.M., Koshler T.M.,
RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria.";
RL Nature 423:81-86 (2003).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581;

```

```
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA Fraser C.M.,
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Stearne;
RA Bretin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017036; AAP27817.1; -
DR EMBL; AE017334; AAT33212.2; -
DR EMBL; AE017225; AAT56105.1; -
DR TIGR; BA4092; -
DR GO; GO:0008462; F:endorpeptidase Clp activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001907; Pept_S14_ClpP.
DR Pfam; PF00574; CLP protease; I.
DR PRINTS; PR00127; CLPPTOTASEP.
KW Protease.
SQ SEQUENCE 251 AA; 28154 MW; 373E72DB6082EE71 CRC64;

Query Match 34.2%; Score 53; DB 2; Length 251;
Best Local Similarity 35.7%; Pred. No. 39;
Matches 10; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ILTSPWTTSSGLWPRLQKAAEFKQLNQ 28
Db 118 MIHNWYASGNADQLRKADDERIQ 145

RESULT 11
AAT33212 PRELIMINARY; PRT; 251 AA.
AC AAT33212;
DT 01-JUN-2004 (TrEMBLrel. 27, Created)
DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE Prophage LambdaBa02, Clp protease family protein.
GN GBAA4092.
OS Bacillus anthracis str. Ames 0581.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017334; AAT33212.2; -
KW Protease.
SQ SEQUENCE 251 AA; 28154 MW; 373E72DB6082EE71 CRC64;

Query Match 34.2%; Score 53; DB 2; Length 251;
Best Local Similarity 35.7%; Pred. No. 39;
Matches 10; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ILTSPWTTSSGLWPRLQKAAEFKQLNQ 28
Db 118 MIHNWYASGNADQLRKADDERIQ 145

RESULT 12
Q9LDZ9 PRELIMINARY; PRT; 460 AA.
AC Q9LDZ9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
```

```
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
DE clone.P0699D11 (Similar to Arabidopsis thaliana chromosome 1 YAC
DE YUP8H12R sequence).
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
RA Wu J., Niimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
RA Ikano M., Itoh S., Itoh T., Itoh Y., Iwabuchi A., Kamiya K.,
RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA Machita K., Machara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
RA Namiki N., Negishi M., Ohta I., Ono N., Sai S., Sakai, K., Shibata M.,
RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
RA Yano M., Jiang J., Gojobori T.;
RT "The genome sequence and structure of rice chromosome 1.";
RL Nature 420:312-316(2002).
DR EMBL; AP002817; BAB03452.1; -
DR EMBL; AP001366; BAA92409.1; -
DR EMBL; AP001366; BAA92405.1; -
DR Gramene; O9LDZ9; -
DR InterPro; IPR003480; Transferase.
DR Pfam; PF02458; transferase; I.
DR F02458; transferase; I.
SQ SEQUENCE 460 AA; 48759 MW; 33BA1150391890B CRC64;

Query Match 34.2%; Score 53; DB 2; Length 460;
Best Local Similarity 43.5%; Pred. No. 73;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 6 WTTSSGLWPRLQKAAEFKQLNQ 28
Db 408 WIASALLWFLAAALSDERLGR 430

RESULT 13
Q9U758 PRELIMINARY; PRT; 280 AA.
AC Q9U758;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein precursor.
GN OrderedLocustNames=SYNW1128;
OS Synechococcus sp. (strain WH8102).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=84588;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825697; PubMed=12917641; DOI=10.1038/nature01943;
RA Palenik B., Brahamsha B., Larimer F.W., Land M.L., Hauser L.,
RA Chain P., Lamerdin J.E., Regala W., Allen E.E., McCaren J.,
RA Paulsen I.T., Dufrene A., Partensky F., Webb E.A., Waterbury J.;
RT "The genome of a motile marine Synechococcus.";
RL Nature 424:1037-1042(2003).
DR EMBL; BX569692; CAE07643.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR000620; DUF6.
DR Pfam; PF00892; DUF6; I.
KW Complete proteome; Hypothetical protein; Signal.
FT SIGNAL 1 22 Potential.
SQ SEQUENCE 280 AA; 28619 MW; B874CA4D098CA7D3 CRC64;

Query Match 33.9%; Score 52.5; DB 2; Length 280;
```

```

RA      Utterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
RA      Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
RA      Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA      Bender C.L., White O., Fraser C.M., Collier A.,
RA      "The complete genome sequence of the Arabidopsis and tomato pathogen
RA      Pseudomonas syringae pv. tomato DC3000."
RA      Proc Natl Acad Sci U S A. 100:10181-10186(2003).
CC      -!- SIMILARITY: Contains 1 HTH lysR-type DNA-binding domain.
EMBL;   AE016956; AAC053670.1; -.
DR      TIGR; PSPT00116; -.
DR      GO; GO:0003700; F:transcription factor activity; IEA.
DR      GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR      InterPro; IPR000847; HTH_LysR.
DR      InterPro; IPR005119; LysR_subst.
DR      InterPro; IPR009058; wing_hlx_DNA_bnd.
DR      Pfam; PF00126; HTH_1; 1.
DR      Pfam; PF03466; LysR_substrate; 1.
DR      PRINTS; PR00039; HTHLYSR.
DR      PROSITE; PSS0931; HTH_LYSR; 1.
DR      Complete proteome; DNA-binding: Transcription regulation.
SQ      SEQUENCE 306 AA; 34681 MW; B598F966404F962F CRC64;
Query Match          33.2%; Score 51.5; DB 2; Length 306;
Best Local Similarity 36.4%; Pred. No. 78;
Matches 12; Conservative       7; Mismatches 7; Indels 7; Gaps
QY      2 LTSPFWTTS--SG-----LWPLQKAAEAFPKOLN 27
DB      255 LSLPWRVTFVSNGNYLVWFRTPGAERLRLS 287
                                     |||||::|::|::|::|
Search completed: November 10, 2004, 13:38:37
Job time : 37.411 secs

```


; ORGANISM: Drosophila melanogaster
US-09-270-767-48003

Query Match 32.4%; Score 57; DB 4; Length 312;
Best Local Similarity 54.2%; Pred. No. 2.9;
Matches 13; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

Qy 4 QPRLNQOQHLPALPIWFLQWL 27
Db 185 QPTLLNTIQHPALHKLWLDNYL 208

RESULT 3

US-09-270-767-48003
; Sequence 48003, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1998-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 48003

; LENGTH: 312

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-270-767-48003

Query Match 32.4%; Score 57; DB 4; Length 312;
Best Local Similarity 54.2%; Pred. No. 2.9;
Matches 13; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

Qy 4 QPRLNQOQHLPALPIWFLQWL 27
Db 185 QPTLLNTIQHPALHKLWLDNYL 208

RESULT 4

US-09-252-991A-22121
; Sequence 22121, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 22121

; LENGTH: 339

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-22121

Query Match 30.4%; Score 53.5; DB 4; Length 339;
Best Local Similarity 48.5%; Pred. No. 9.9;
Matches 16; Conservative 1; Mismatches 5; Indels 11; Gaps 3;

Qy 3 LQPRLLNQOQHLPALPIWFLQWLRLHP 31
Db 258 VQPR-RQPHLHLPAPARQLP-----WLRHLP 283

RESULT 5

US-09-543-681A-5973

; Sequence 5973, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; FILE REFERENCE: 2709-1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 5973

; LENGTH: 398

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-5973

Query Match 30.1%; Score 53; DB 4; Length 398;

Best Local Similarity 37.5%; Pred. No. 14;

Matches 12; Conservative 6; Mismatches 8; Indels 6; Gaps 1;

Qy 1 RTLQPRLLNQOQHLPALPIWFLQWLRLHPL 32
Db 258 RFLAVRMKQHEQSGLE-----IARWLKQHPL 283

RESULT 6

US-09-252-991A-31679

; Sequence 31679, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 31679

; LENGTH: 199

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31679

Query Match 29.5%; Score 52; DB 4; Length 199;

Best Local Similarity 41.7%; Pred. No. 8.8;

Matches 15; Conservative 4; Mismatches 7; Indels 10; Gaps 3;

Qy 6 RLLNQOQHLPALPIWF-----LIQW-----LRLHPL 32
Db 26 QLVFFQLEHFPILQRFATEKELVQWAGPALR-HPL 60

RESULT 7

US-09-489-039A-12472

; Sequence 12472, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 12472

; LENGTH: 301

```
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12472

Query Match      29.3%; Score 51.5; DB 4; Length 301;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 2; Mismatches 9; Indels 7; Gaps 1;

QY      6 RLQOQQOHLPALPIWFLQWL-----RLH 30
DB      76 RLQOQLEPLIVFVILLGLLCIGAGKVRLLH 107

RESULT 8
US-09-335-409-5
; Sequence 5, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-5

Query Match      29.3%; Score 51.5; DB 3; Length 7257;
Best Local Similarity 41.7%; Pred. No. 6.2e+02;
Matches 15; Conservative 4; Mismatches 12; Indels 5; Gaps 2;

QY      2 TLQPRLLNQOQHL-PA-----LPIWFLQWLRLHPL 32
DB      4499 TVDARLVVRQAGLTTPAQATVPVAFLTAWLALHDL 4534

RESULT 9
US-09-568-102-5
; Sequence 5, Application US/09568102
; Patent No. 6348404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-5

Query Match      29.3%; Score 51.5; DB 3; Length 7257;
Best Local Similarity 41.7%; Pred. No. 6.2e+02;
Matches 15; Conservative 4; Mismatches 12; Indels 5; Gaps 2;
```

```
QY      2 TLQPRLLNQOQHL-PA-----LPIWFLQWLRLHPL 32
DB      4499 TVDARLVVRQAGLTTPAQATVPVAFLTAWLALHDL 4534

RESULT 10
US-09-567-969-5
; Sequence 5, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-969-5

Query Match      29.3%; Score 51.5; DB 3; Length 7257;
Best Local Similarity 41.7%; Pred. No. 6.2e+02;
Matches 15; Conservative 4; Mismatches 12; Indels 5; Gaps 2;

QY      2 TLQPRLLNQOQHL-PA-----LPIWFLQWLRLHPL 32
DB      4499 TVDARLVVRQAGLTTPAQATVPVAFLTAWLALHDL 4534

RESULT 11
US-09-568-480-5
; Sequence 5, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-5

Query Match      29.3%; Score 51.5; DB 3; Length 7257;
Best Local Similarity 41.7%; Pred. No. 6.2e+02;
Matches 15; Conservative 4; Mismatches 12; Indels 5; Gaps 2;

QY      2 TLQPRLLNQOQHL-PA-----LPIWFLQWLRLHPL 32
DB      4499 TVDARLVVRQAGLTTPAQATVPVAFLTAWLALHDL 4534
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 34.3765 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-37

Perfect score: 176

Sequence: 1 RTLQRLNQOQHLPALPIWFLQWLRLHPL 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	176	100.0	32	14	US-10-092-750-37 Sequence 37, Appl
2	65.5	37.2	421	17	US-10-425-115-271834 Sequence 271834,
3	57.5	32.7	567	15	US-10-037-417-106 Sequence 106, Appl
4	57	32.4	403	15	US-10-037-417-106 Sequence 4, Appl
5	57	32.4	403	15	US-10-220-475A-4 Sequence 4, Appl
6	57	32.4	403	15	US-10-419-629-4 Sequence 4, Appl
7	57	32.4	403	15	US-10-072-012-448 Sequence 4, Appl
8	57	32.4	403	16	US-10-696-699A-4 Sequence 4, Appl
9	56.5	32.1	71	15	US-10-425-114-71590 Sequence 71590, A
10	56.5	32.1	400	15	US-10-220-475A-17 Sequence 17, Appl
11	56.5	32.1	400	15	US-10-072-012-451 Sequence 451, Appl
12	56.5	32.1	403	15	US-10-220-475A-2 Sequence 2, Appl
13	56.5	32.1	403	15	US-10-072-012-449 Sequence 449, Appl
	56.5	32.1	566	10	US-09-775-803-12 Sequence 12, Appl

14	56.5	32.1	567	15	US-10-037-417-105 Sequence 105, Appl
15	56	31.8	90	15	US-10-424-599-279796 Sequence 279796,
16	55	31.2	167	10	US-09-764-891-51116 Sequence 51116, Ap
17	54.5	31.0	170	17	US-10-425-115-362182 Sequence 362182,
18	54.5	31.0	131	17	US-10-425-115-324612 Sequence 324612,
19	54.5	31.0	312	14	US-10-300-846-18 Sequence 18, Appl
20	54.5	31.0	406	15	US-10-220-475A-18 Sequence 18, Appl
21	54.5	31.0	406	15	US-10-072-012-447 Sequence 447, Appl
22	54	30.7	79	16	US-10-437-963-197766 Sequence 197766,
23	53.5	30.4	68	16	US-10-767-701-36749 Sequence 36749, A
24	53.5	30.4	314	9	US-09-771-209-78 Sequence 78, Appl
25	53.5	30.4	416	15	US-10-072-012-124 Sequence 124, Appl
26	53	30.1	64	9	US-09-864-761-37217 Sequence 37217, A
27	53	30.1	67	16	US-10-437-963-158507 Sequence 158507,
28	53	30.1	290	17	US-10-739-930-10228 Sequence 10228, A
29	53	30.1	2430	15	US-10-320-797-3361 Sequence 3361, Ap
30	52.5	29.8	97	11	US-09-864-408A-3904 Sequence 3904, Ap
31	52.5	29.8	403	15	US-10-072-012-450 Sequence 450, Appl
32	52.5	29.8	896	16	US-10-437-963-136508 Sequence 136508,
33	52	29.5	98	17	US-10-425-115-330208 Sequence 330208,
34	52	29.5	132	17	US-10-425-115-240589 Sequence 240589,
35	52	29.5	156	16	US-10-767-701-60525 Sequence 60525, A
36	52	29.5	186	15	US-10-424-599-270451 Sequence 270451,
37	52	29.5	266	14	US-10-017-161-2026 Sequence 1672, Ap
38	52	29.5	266	14	US-10-292-798-1672 Sequence 1672, Ap
39	52	29.5	390	15	US-10-424-599-270450 Sequence 270450,
40	51.5	29.3	267	17	US-10-425-115-240781 Sequence 240781,
41	51.5	29.3	314	9	US-09-771-209-80 Sequence 80, Appl
42	51.5	29.3	314	9	US-09-771-730-41 Sequence 41, Appl
43	51.5	29.3	314	10	US-09-908-006A-51 Sequence 51, Appl
44	51.5	29.3	7257	13	US-10-014-717-5 Sequence 5, Appl
45	51	29.0	216	9	US-09-747-155-373 Sequence 373, Appl

ALIGNMENTS

RESULT 1

US-10-092-750-37
; Sequence 37, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-37

Query Match 100.0%; Score 176; DB 14; Length 32;

Best Local Similarity 100.0%; Pred. No. 2.8e-15; Indels 0; Gaps 0;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2

US-10-425-115-271834
; Sequence 271834, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 271834
LENGTH: 421
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_179504C.1.pep
US-10-425-115-271834

Query Match 37.2%; Score 65.5; DB 17; Length 421;
Best Local Similarity 48.4%; Pred. No. 4;
Matches 15; Conservative 1; Mismatches 6; Indels 9; Gaps 2;

QY 9 QNQQOHLPALPIPF-----LLQWLRLHP 31
DB 369 EQQPHQFQLP-NFLQRQPPVLLPWLRLHP 398

RESULT 3
US-10-037-417-106
Sequence 106, Application US/10037417
Publication No. US20040052806A1
GENERAL INFORMATION:
APPLICANT: Kexuda, Ramesh
APPLICANT: Alsobrook II, John P
APPLICANT: Tchernev, Velizar T
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly A
APPLICANT: Patturajan, Meera
APPLICANT: Grosse, William M
APPLICANT: Lepley, Denise M
APPLICANT: Burgess, Catherine E
APPLICANT: Vernet, Corine A.M.
APPLICANT: Li, Li
APPLICANT: Gorman, Linda
APPLICANT: Edinger, Shlomit R
APPLICANT: Sciore, Paul
APPLICANT: Ellerman, Karen
APPLICANT: Malyankar, Uriel M
APPLICANT: Rothenberg, Mark
APPLICANT: Stone, David J
APPLICANT: Boldog, Ferenc L
APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Surash G
APPLICANT: Anderson, David W
APPLICANT: Padigaru, Muralidhara
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Miller, Charles E
APPLICANT: Eisen, Andrew J
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-235
CURRENT APPLICATION NUMBER: US/10/037,417
CURRENT FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: 60/260,018
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,360
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/272,411
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/272,817
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/291,186
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/303,231

PRIOR FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/305,060
PRIOR FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 60/318,405
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/318,700
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 227
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 106
LENGTH: 567
TYPE: PRT
ORGANISM: Mus musculus
US-10-037-417-106

Query Match 32.7%; Score 57.5; DB 15; Length 567;
Best Local Similarity 45.7%; Pred. No. 55;
Matches 14; Conservative 2; Mismatches 11; Indels 3; Gaps 1;

QY 5 PRLNQOQHLPAL---PIWFLQWLRLHP 31
DB 411 PQLTQVLGHNPWLCDCGLWRFLQWLRLHP 440

RESULT 4
US-10-220-475A-4
Sequence 4, Application US/10220475A
Publication No. US20040023227A1
GENERAL INFORMATION:
APPLICANT: INOUE, KEIZO
APPLICANT: ARAI, HIROYUKI
APPLICANT: ARITA, MAKOTO
TITLE OF INVENTION: DNA ENCODING SQUALENE EPOXIDASE-PROMOTING FACTOR
FILE REFERENCE: 04853.0096-00000
CURRENT APPLICATION NUMBER: US/10/220,475A
CURRENT FILING DATE: 2002-08-30
PRIOR APPLICATION NUMBER: PCT/JP01/01592
PRIOR FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: JP 2000-57743
PRIOR FILING DATE: 2000-03-02
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 403
TYPE: PRT
ORGANISM: Homo sapiens
US-10-220-475A-4

Query Match 32.4%; Score 57; DB 15; Length 403;
Best Local Similarity 44.4%; Pred. No. 46;
Matches 16; Conservative 3; Mismatches 7; Indels 10; Gaps 2;

QY 3 LQPR-----LLNQOQHLPALP---IWFLQWLRL 28
DB 8 LSPRKEALAKFRENQDVLPAFPNDDYFLRLRL 43

RESULT 5
US-10-419-629-4
Sequence 4, Application US/10419629
Publication No. US20040023915A1
GENERAL INFORMATION:
APPLICANT: Sanders, Bob G.
APPLICANT: Kline, Kimberly
APPLICANT: Yu, Weiping
APPLICANT: Liu, Hui
APPLICANT: Hantash, Feras
TITLE OF INVENTION: Tocopherol Associated Protein and Uses Thereof
FILE REFERENCE: D8453
CURRENT APPLICATION NUMBER: US/10/419,629
CURRENT FILING DATE: 2003-04-21
PRIOR APPLICATION NUMBER: US 60/373,870
PRIOR FILING DATE: 2002-04-19

```

; SEQ ID NO 71590
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:

```

; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.

; OTHER INFORMATION: Clone ID: UC-ZMFLMO17164H05_FLI.pap
US-10-425-114-71590

Query Match 32.1%; Score 56.5; DB 15; Length 71;
Best Local Similarity 32.5%; Pred. No. 9.9;
Matches 13; Conservative 9; Mismatches 7; Indels 11; Gaps 2;

QY 4 QPRLNQOQHLPALPI-----WFLQWL-----RLHPL 32
DB 18 EPRIVSKQELPQYTLQRFQYNSRLEHLSDRSFVHPL 57

RESULT 9

US-10-220-475A-17
; Sequence 17, Application US/10220475A
; Publication No. US20040023227A1
; GENERAL INFORMATION:
; APPLICANT: INOUE, KEIZO
; APPLICANT: ARAI, HIROYUKI
; APPLICANT: ARITA, MAKOTO
; TITLE OF INVENTION: DNA ENCODING SQUALENE EPOXIDASE-PROMOTING FACTOR
; FILE REFERENCE: 04853.0096-00000
; CURRENT APPLICATION NUMBER: US/10/220.475A
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: PCT/JP01/01592
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: JP 2000-57743
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-220-475A-17

Query Match 32.1%; Score 56.5; DB 15; Length 400;
Best Local Similarity 56.5%; Pred. No. 53;
Matches 13; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

QY 9 ONQOQHLPALP---IWFLQWL 28
DB 21 ENVQDVLFPALPNPDYFLLRWLR 43

RESULT 10

US-10-072-012-451
; Sequence 451, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072.012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 451
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-072-012-451

Query Match 32.1%; Score 56.5; DB 15; Length 400;
Best Local Similarity 56.5%; Pred. No. 53;
Matches 13; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

QY 9 ONQOQHLPALP---IWFLQWL 28
DB 21 ENVQDVLFPALPNPDYFLLRWLR 43

RESULT 11

US-10-220-475A-2
; Sequence 2, Application US/10220475A
; Publication No. US20040023227A1
; GENERAL INFORMATION:
; APPLICANT: INOUE, KEIZO
; APPLICANT: ARAI, HIROYUKI
; APPLICANT: ARITA, MAKOTO
; TITLE OF INVENTION: DNA ENCODING SQUALENE EPOXIDASE-PROMOTING FACTOR
; FILE REFERENCE: 04853.0096-00000
; CURRENT APPLICATION NUMBER: US/10/220.475A
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: PCT/JP01/01592
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: JP 2000-57743
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-220-475A-2

Query Match 32.1%; Score 56.5; DB 15; Length 403;
Best Local Similarity 56.5%; Pred. No. 53;
Matches 13; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

QY 9 ONQOQHLPALP---IWFLQWL 28
DB 21 ENVQDVLFPALPNPDYFLLRWLR 43

RESULT 12

US-10-072-012-449
; Sequence 449, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerkhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072.012
; CURRENT FILING DATE: 2002-01-31
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 449
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-072-012-449

Query Match 32.1%; Score 56.5; DB 15; Length 403;
Best Local Similarity 56.5%; Pred. No. 53;
Matches 13; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

Qy 9 QNQQQHLPALP---IWFLQWLRL 28
:|||||:|||||:
Db 21 ENVDVLPALPNPDYELLRL 43

RESULT 13

US-09-775-803-12
; Sequence 12, Application US/09775803
; Publication No. US20030167487A1
; GENERAL INFORMATION:
; APPLICANT: COR Therapeutics, Inc.
; APPLICANT: Ramakrishnan, Vanitha
; APPLICANT: Phillips, David
; TITLE OF INVENTION: Transgenic Animals Having a Modified Glycoprotein V
; FILE REFERENCE: 44481-5044-US
; CURRENT APPLICATION NUMBER: US/09/775,803
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/17594
; PRIOR FILING DATE: 1999-08-04
; PRIOR APPLICATION NUMBER: US 60/109,797
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-775-803-12

Query Match 32.1%; Score 56.5; DB 10; Length 566;
Best Local Similarity 46.7%; Pred. No. 74;
Matches 14; Conservative 2; Mismatches 11; Indels 3; Gaps 1;

Qy 5 PRLNQQQHLPAL---PIWFLLQWLRLHP 31
:|||||:|||||:
Db 411 PQLTQVLGHNPWLCDCGLWPFLQWLRLHHP 440

RESULT 14

US-10-037-417-105
; Sequence 105, Application US/10037417
; Publication No. US20040052806A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Anderson, David W
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Miller, Charles E
; APPLICANT: Eisen, Andrew J
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-235
; CURRENT APPLICATION NUMBER: US/10/037,417
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/260,018
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: 60/260,360
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/272,411

```
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/272,817
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/291,186
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/303,231
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/305,060
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/318,405
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/318,700
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 227
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 105
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-417-105

Query Match      32.1%; Score 56.5; DB 15; Length 567;
Best Local Similarity 46.7%; Pred. No. 74;
Matches 14; Conservative 2; Mismatches 11; Indels 3; Gaps 1;

QY 5 PRLNQOQHLPAL---PIWFLLQWLRLHP 31
DB 411 PQLTQVLLGHNPWLCDCGLMPFLQWLRLHP 440

RESULT 15
US-10-424-599-279796
; Sequence 279796, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 279796
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(90)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_94679C.1.pep
US-10-424-599-279796

Query Match      31.8%; Score 56; DB 15; Length 90;
Best Local Similarity 38.5%; Pred. No. 14;
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 3 LQPRLLNQOQHLPALPIWFLLQWLRL 28
DB 17 MKPRIVSLKIEHLVXLPRWFTQVLK 42

Search completed: November 11, 2004, 01:28:19
Job time : 35.4265 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 7.12953 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-37

Perfect score: 176

Sequence: 1 RTLQRLNQOQHLPALPIWFLQWLRHPL 32

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	38.1	1360	2 T33922	hypothetical prote
2	60	34.1	225	2 AC1896	hypothetical prote
3	55	31.2	152	2 F82493	hypothetical prote
4	55	31.2	1421	2 T00333	hypothetical prote
5	53.5	30.4	314	2 H23701	olfactory receptor
6	53.5	30.4	314	2 S29707	olfactory receptor
7	53	30.1	403	2 JC7708	tocopherol associa
8	52.5	29.8	448	2 T31334	hglD protein - Ana
9	52.5	29.8	453	2 AB2475	heterocyst glycoli
10	52	29.5	292	2 B90558	hypothetical prote
11	52	29.5	862	2 T38996	hypothetical prote
12	51.5	29.3	240	2 T40985	hypothetical prote
13	51.5	29.3	314	2 A73286	olfactory receptor
14	51	29.0	219	2 E75219	hypothetical prote
15	51	29.0	313	2 S20571	olfactory receptor
16	51	29.0	448	2 AC3602	transporter, mfs s
17	51	29.0	482	2 T36045	hypothetical prote
18	51	29.0	498	2 AE2167	hypothetical prote
19	51	29.0	833	2 T22139	hypothetical prote
20	50	28.4	180	2 I84689	escapes X-inactiva
21	50	28.4	753	2 AH2543	cation-transportin
22	50	28.4	859	2 D96502	hypothetical prote
23	50	28.4	1024	1 EBEC	beta-galactosidase
24	50	28.4	1024	2 E90678	beta-D-galactosida
25	50	28.4	1024	2 A85529	beta-D-galactosida
26	49.5	28.1	264	2 A23207	22K zein precursor
27	49.5	28.1	266	1 Z1ZM23	22K zein precursor
28	49.5	28.1	266	2 B22831	hypothetical prote
29	49.5	28.1	303	2 AG1925	hypothetical prote

ALIGNMENTS

RESULT 1

T33922
hypothetical protein Y8A9A.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T33922
R:Courtney, L.; Langston, Y.; Drone, K.; Mead, K.
submitted to the EMBL Data Library, February 1999
A:Description: The sequence of C. elegans cosmid Y8A9A.
A:Reference number: Z21439
A:Accession: T33922
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1360 <COU>
A:Cross-references: UNIPROT:Q9YVK4; EMBL:AF125461; PIDN:AAD12852.1; GSPDB:GN000020; CESP:
A:Experimental source: Strain Bristol N2; clone Y8A9A
C:Genetics:
A:Gene: CESP:Y8A9A.2
A:Map position: 2
A:Introns: 43/3; 69/3; 134/3; 174/1; 604/2; 643/3; 718/3; 787/3; 848/2; 944/2; 1044/2; 1104/2; 1204/2; 1304/2; 1404/2; 1504/2; 1604/2; 1704/2; 1804/2; 1904/2; 2004/2; 2104/2; 2204/2; 2304/2; 2404/2; 2504/2; 2604/2; 2704/2; 2804/2; 2904/2; 3004/2; 3104/2; 3204/2; 3304/2; 3404/2; 3504/2; 3604/2; 3704/2; 3804/2; 3904/2; 4004/2; 4104/2; 4204/2; 4304/2; 4404/2; 4504/2; 4604/2; 4704/2; 4804/2; 4904/2; 5004/2; 5104/2; 5204/2; 5304/2; 5404/2; 5504/2; 5604/2; 5704/2; 5804/2; 5904/2; 6004/2; 6104/2; 6204/2; 6304/2; 6404/2; 6504/2; 6604/2; 6704/2; 6804/2; 6904/2; 7004/2; 7104/2; 7204/2; 7304/2; 7404/2; 7504/2; 7604/2; 7704/2; 7804/2; 7904/2; 8004/2; 8104/2; 8204/2; 8304/2; 8404/2; 8504/2; 8604/2; 8704/2; 8804/2; 8904/2; 9004/2; 9104/2; 9204/2; 9304/2; 9404/2; 9504/2; 9604/2; 9704/2; 9804/2; 9904/2; 10004/2; 10104/2; 10204/2; 10304/2; 10404/2; 10504/2; 10604/2; 10704/2; 10804/2; 10904/2; 11004/2; 11104/2; 11204/2; 11304/2; 11404/2; 11504/2; 11604/2; 11704/2; 11804/2; 11904/2; 12004/2; 12104/2; 12204/2; 12304/2; 12404/2; 12504/2; 12604/2; 12704/2; 12804/2; 12904/2; 13004/2; 13104/2; 13204/2; 13304/2; 13404/2; 13504/2; 13604/2; 13704/2; 13804/2; 13904/2; 14004/2; 14104/2; 14204/2; 14304/2; 14404/2; 14504/2; 14604/2; 14704/2; 14804/2; 14904/2; 15004/2; 15104/2; 15204/2; 15304/2; 15404/2; 15504/2; 15604/2; 15704/2; 15804/2; 15904/2; 16004/2; 16104/2; 16204/2; 16304/2; 16404/2; 16504/2; 16604/2; 16704/2; 16804/2; 16904/2; 17004/2; 17104/2; 17204/2; 17304/2; 17404/2; 17504/2; 17604/2; 17704/2; 17804/2; 17904/2; 18004/2; 18104/2; 18204/2; 18304/2; 18404/2; 18504/2; 18604/2; 18704/2; 18804/2; 18904/2; 19004/2; 19104/2; 19204/2; 19304/2; 19404/2; 19504/2; 19604/2; 19704/2; 19804/2; 19904/2; 20004/2; 20104/2; 20204/2; 20304/2; 20404/2; 20504/2; 20604/2; 20704/2; 20804/2; 20904/2; 21004/2; 21104/2; 21204/2; 21304/2; 21404/2; 21504/2; 21604/2; 21704/2; 21804/2; 21904/2; 22004/2; 22104/2; 22204/2; 22304/2; 22404/2; 22504/2; 22604/2; 22704/2; 22804/2; 22904/2; 23004/2; 23104/2; 23204/2; 23304/2; 23404/2; 23504/2; 23604/2; 23704/2; 23804/2; 23904/2; 24004/2; 24104/2; 24204/2; 24304/2; 24404/2; 24504/2; 24604/2; 24704/2; 24804/2; 24904/2; 25004/2; 25104/2; 25204/2; 25304/2; 25404/2; 25504/2; 25604/2; 25704/2; 25804/2; 25904/2; 26004/2; 26104/2; 26204/2; 26304/2; 26404/2; 26504/2; 26604/2; 26704/2; 26804/2; 26904/2; 27004/2; 27104/2; 27204/2; 27304/2; 27404/2; 27504/2; 27604/2; 27704/2; 27804/2; 27904/2; 28004/2; 28104/2; 28204/2; 28304/2; 28404/2; 28504/2; 28604/2; 28704/2; 28804/2; 28904/2; 29004/2; 29104/2; 29204/2; 29304/2; 29404/2; 29504/2; 29604/2; 29704/2; 29804/2; 29904/2; 30004/2; 30104/2; 30204/2; 30304/2; 30404/2; 30504/2; 30604/2; 30704/2; 30804/2; 30904/2; 31004/2; 31104/2; 31204/2; 31304/2; 31404/2; 31504/2; 31604/2; 31704/2; 31804/2; 31904/2; 32004/2; 32104/2; 32204/2; 32304/2; 32404/2; 32504/2; 32604/2; 32704/2; 32804/2; 32904/2; 33004/2; 33104/2; 33204/2; 33304/2; 33404/2; 33504/2; 33604/2; 33704/2; 33804/2; 33904/2; 34004/2; 34104/2; 34204/2; 34304/2; 34404/2; 34504/2; 34604/2; 34704/2; 34804/2; 34904/2; 35004/2; 35104/2; 35204/2; 35304/2; 35404/2; 35504/2; 35604/2; 35704/2; 35804/2; 35904/2; 36004/2; 36104/2; 36204/2; 36304/2; 36404/2; 36504/2; 36604/2; 36704/2; 36804/2; 36904/2; 37004/2; 37104/2; 37204/2; 37304/2; 37404/2; 37504/2; 37604/2; 37704/2; 37804/2; 37904/2; 38004/2; 38104/2; 38204/2; 38304/2; 38404/2; 38504/2; 38604/2; 38704/2; 38804/2; 38904/2; 39004/2; 39104/2; 39204/2; 39304/2; 39404/2; 39504/2; 39604/2; 39704/2; 39804/2; 39904/2; 40004/2; 40104/2; 40204/2; 40304/2; 40404/2; 40504/2; 40604/2; 40704/2; 40804/2; 40904/2; 41004/2; 41104/2; 41204/2; 41304/2; 41404/2; 41504/2; 41604/2; 41704/2; 41804/2; 41904/2; 42004/2; 42104/2; 42204/2; 42304/2; 42404/2; 42504/2; 42604/2; 42704/2; 42804/2; 42904/2; 43004/2; 43104/2; 43204/2; 43304/2; 43404/2; 43504/2; 43604/2; 43704/2; 43804/2; 43904/2; 44004/2; 44104/2; 44204/2; 44304/2; 44404/2; 44504/2; 44604/2; 44704/2; 44804/2; 44904/2; 45004/2; 45104/2; 45204/2; 45304/2; 45404/2; 45504/2; 45604/2; 45704/2; 45804/2; 45904/2; 46004/2; 46104/2; 46204/2; 46304/2; 46404/2; 46504/2; 46604/2; 46704/2; 46804/2; 46904/2; 47004/2; 47104/2; 47204/2; 47304/2; 47404/2; 47504/2; 47604/2; 47704/2; 47804/2; 47904/2; 48004/2; 48104/2; 48204/2; 48304/2; 48404/2; 48504/2; 48604/2; 48704/2; 48804/2; 48904/2; 49004/2; 49104/2; 49204/2; 49304/2; 49404/2; 49504/2; 49604/2; 49704/2; 49804/2; 49904/2; 50004/2; 50104/2; 50204/2; 50304/2; 50404/2; 50504/2; 50604/2; 50704/2; 50804/2; 50904/2; 51004/2; 51104/2; 51204/2; 51304/2; 51404/2; 51504/2; 51604/2; 51704/2; 51804/2; 51904/2; 52004/2; 52104/2; 52204/2; 52304/2; 52404/2; 52504/2; 52604/2; 52704/2; 52804/2; 52904/2; 53004/2; 53104/2; 53204/2; 53304/2; 53404/2; 53504/2; 53604/2; 53704/2; 53804/2; 53904/2; 54004/2; 54104/2; 54204/2; 54304/2; 54404/2; 54504/2; 54604/2; 54704/2; 54804/2; 54904/2; 55004/2; 55104/2; 55204/2; 55304/2; 55404/2; 55504/2; 55604/2; 55704/2; 55804/2; 55904/2; 56004/2; 56104/2; 56204/2; 56304/2; 56404/2; 56504/2; 56604/2; 56704/2; 56804/2; 56904/2; 57004/2; 57104/2; 57204/2; 57304/2; 57404/2; 57504/2; 57604/2; 57704/2; 57804/2; 57904/2; 58004/2; 58104/2; 58204/2; 58304/2; 58404/2; 58504/2; 58604/2; 58704/2; 58804/2; 58904/2; 59004/2; 59104/2; 59204/2; 59304/2; 59404/2; 59504/2; 59604/2; 59704/2; 59804/2; 59904/2; 60004/2; 60104/2; 60204/2; 60304/2; 60404/2; 60504/2; 60604/2; 60704/2; 60804/2; 60904/2; 61004/2; 61104/2; 61204/2; 61304/2; 61404/2; 61504/2; 61604/2; 61704/2; 61804/2; 61904/2; 62004/2; 62104/2; 62204/2; 62304/2; 62404/2; 62504/2; 62604/2; 62704/2; 62804/2; 62904/2; 63004/2; 63104/2; 63204/2; 63304/2; 63404/2; 63504/2; 63604/2; 63704/2; 63804/2; 63904/2; 64004/2; 64104/2; 64204/2; 64304/2; 64404/2; 64504/2; 64604/2; 64704/2; 64804/2; 64904/2; 65004/2; 65104/2; 65204/2; 65304/2; 65404/2; 65504/2; 65604/2; 65704/2; 65804/2; 65904/2; 66004/2; 66104/2; 66204/2; 66304/2; 66404/2; 66504/2; 66604/2; 66704/2; 66804/2; 66904/2; 67004/2; 67104/2; 67204/2; 67304/2; 67404/2; 67504/2; 67604/2; 67704/2; 67804/2; 67904/2; 68004/2; 68104/2; 68204/2; 68304/2; 68404/2; 68504/2; 68604/2; 68704/2; 68804/2; 68904/2; 69004/2; 69104/2; 69204/2; 69304/2; 69404/2; 69504/2; 69604/2; 69704/2; 69804/2; 69904/2; 70004/2; 70104/2; 70204/2; 70304/2; 70404/2; 70504/2; 70604/2; 70704/2; 70804/2; 70904/2; 71004/2; 71104/2; 71204/2; 71304/2; 71404/2; 71504/2; 71604/2; 71704/2; 71804/2; 71904/2; 72004/2; 72104/2; 72204/2; 72304/2; 72404/2; 72504/2; 72604/2; 72704/2; 72804/2; 72904/2; 73004/2; 73104/2; 73204/2; 73304/2; 73404/2; 73504/2; 73604/2; 73704/2; 73804/2; 73904/2; 74004/2; 74104/2; 74204/2; 74304/2; 74404/2; 74504/2; 74604/2; 74704/2; 74804/2; 74904/2; 75004/2; 75104/2; 75204/2; 75304/2; 75404/2; 75504/2; 75604/2; 75704/2; 75804/2; 75904/2; 76004/2; 76104/2; 76204/2; 76304/2; 76404/2; 76504/2; 76604/2; 76704/2; 76804/2; 76904/2; 77004/2; 77104/2; 77204/2; 77304/2; 77404/2; 77504/2; 77604/2; 77704/2; 77804/2; 77904/2; 78004/2; 78104/2; 78204/2; 78304/2; 78404/2; 78504/2; 78604/2; 78704/2; 78804/2; 78904/2; 79004/2; 79104/2; 79204/2; 79304/2; 79404/2; 79504/2; 79604/2; 79704/2; 79804/2; 79904/2; 80004/2; 80104/2; 80204/2; 80304/2; 80404/2; 80504/2; 80604/2; 80704/2; 80804/2; 80904/2; 81004/2; 81104/2; 81204/2; 81304/2; 81404/2; 81504/2; 81604/2; 81704/2; 81804/2; 81904/2; 82004/2; 82104/2; 82204/2; 82304/2; 82404/2; 82504/2; 82604/2; 82704/2; 82804/2; 82904/2; 83004/2; 83104/2; 83204/2; 83304/2; 83404/2; 83504/2; 83604/2; 83704/2; 83804/2; 83904/2; 84004/2; 84104/2; 84204/2; 84304/2; 84404/2; 84504/2; 84604/2; 84704/2; 84804/2; 84904/2; 85004/2; 85104/2; 85204/2; 85304/2; 85404/2; 85504/2; 85604/2; 85704/2; 85804/2; 85904/2; 86004/2; 86104/2; 86204/2; 86304/2; 86404/2; 86504/2; 86604/2; 86704/2; 86804/2; 86904/2; 87004/2; 87104/2; 87204/2; 87304/2; 87404/2; 87504/2; 87604/2; 87704/2; 87804/2; 87904/2; 88004/2; 88104/2; 88204/2; 88304/2; 88404/2; 88504/2; 88604/2; 88704/2; 88804/2; 88904/2; 89004/2; 89104/2; 89204/2; 89304/2; 89404/2; 89504/2; 89604/2; 89704/2; 89804/2; 89904/2; 90004/2; 90104/2; 90204/2; 90304/2; 90404/2; 90504/2; 90604/2; 90704/2; 90804/2; 90904/2; 91004/2; 91104/2; 91204/2; 91304/2; 91404/2; 91504/2; 91604/2; 91704/2; 91804/2; 91904/2; 92004/2; 92104/2; 92204/2; 92304/2; 92404/2; 92504/2; 92604/2; 92704/2; 92804/2; 92904/2; 93004/2; 93104/2; 93204/2; 93304/2; 93404/2; 93504/2; 93604/2; 93704/2; 93804/2; 93904/2; 94004/2; 94104/2; 94204/2; 94304/2; 94404/2; 94504/2; 94604/2; 94704/2; 94804/2; 94904/2; 95004/2; 95104/2; 95204/2; 95304/2; 95404/2; 95504/2; 95604/2; 95704/2; 95804/2; 95904/2; 96004/2; 96104/2; 96204/2; 96304/2; 96404/2; 96504/2; 96604/2; 96704/2; 96804/2; 96904/2; 97004/2; 97104/2; 97204/2; 97304/2; 97404/2; 97504/2; 97604/2; 97704/2; 97804/2; 97904/2; 98004/2; 98104/2; 98204/2; 98304/2; 98404/2; 98504/2; 98604/2; 98704/2; 98804/2; 98904/2; 99004/2; 99104/2; 99204/2; 99304/2; 99404/2; 99504/2; 99604/2; 99704/2; 99804/2; 99904/2; 100004/2; 100014/2; 100024/2; 100034/2; 100044/2; 100054/2; 100064/2; 1000

```

Best Local Similarity 45.5%; Pred. No. 1.1;
Matches 15; Conservative 3; Mismatches 5; Indels 10; Gaps 2;

Qy 6 RLLQNOQHLP-----ALPIWFLQWLRLHPL 32
Db 88 RGLQNSCHLPVVPVTAALPLWL----LRLHSI 116

RESULT 3
H2493
hypothetical protein VCA0173 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Species: Vibrio cholerae
C:Accession: F82493
R:Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: F82493
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-162 <HEI>
A:Cross-references: UNIPROT:Q9RMZ3; GB:AE004357; GB:AE003853; NID:g9657547; PIDN:AAF9608
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0173
A:Map position: 2

Query Match 31.2%; Score 55; DB 2; Length 162;
Best Local Similarity 69.2%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 14 HLPALPIWFLQWL 26
Db 15 HLPPLPSWFFLAW 27

RESULT 4
T00333
hypothetical protein KIAA0560 - human
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
C:Accession: T00333
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14066; MUID:98290545; PMID:9628581
A:Accession: T00333
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1421 <NAG>
A:Cross-references: UNIPROT:O60306; EMBL:AB011132; NID:dl185402; PIDN:BAA25486.1
A:Experimental source: brain; clone HH1648
C:Genetics:
A:Note: KIAA0560

Query Match 31.2%; Score 55; DB 2; Length 1421;
Best Local Similarity 42.9%; Pred. No. 31;
Matches 12; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 2 TLQPLLQNOQHLPALPIWFLQWLRL 29
Db 92 SLEVDLRSVQQQLISLPWMKGLQLARL 119

RESULT 5
H23701
olfactory receptor I9 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 09-Jul-2004
C:Accession: H23701

R:Buck, L.; Axel, R.
Cell 65, 175-187, 1991
A:Title: A novel multigene family may encode odorant receptors: a molecular basis for od
A:Reference number: A23701; MUID:91191556; PMID:1840504
A:Accession: H23701
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-314 <BUC>
A:Cross-references: UNIPROT:P23272; GB:M64388; NID:g205837; PIDN:AAA41751.1; PID:g205838
A:Superfamily: olfactory receptor OR14
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 30.4%; Score 53.5; DB 2; Length 314;
Best Local Similarity 40.7%; Pred. No. 11;
Matches 11; Conservative 5; Mismatches 4; Indels 7; Gaps 1;

Qy 5 PRLQNOQHLPALP-----IWFL 24
Db 79 PKLLQNMQSQVPSIPYAGCLAQIYFFL 105

RESULT 6
S29707
olfactory receptor OR5 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S29707; B37286
R:Raming, K.; Krieger, J.; Strotmann, J.; Boekhoff, I.; Kubick, S.; Baumstark, C.; Breer,
Nature 361, 353-356, 1993
A:Title: Cloning and expression of odorant receptors.
A:Reference number: S29707; MUID:93149273; PMID:7678922
A:Accession: S29707
A:Molecule type: mRNA
A:Residues: 1-314 <RAM>
A:Cross-references: UNIPROT:Q04059
R:Buck, L.; Axel, R.
Cell 65, 175-187, 1991
A:Title: A novel multigene family may encode odorant receptors: a molecular basis for od
A:Reference number: A23701; MUID:91191556; PMID:1840504
A:Accession: B37286
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 193-236 <BUC>
A:Cross-references: GB:M64375; NID:g205811; PIDN:AAA41738.1; PID:g205812
C:Superfamily: olfactory receptor OR14
C:Keywords: G protein-coupled receptor; membrane protein

Query Match 30.4%; Score 53.5; DB 2; Length 314;
Best Local Similarity 40.7%; Pred. No. 11;
Matches 11; Conservative 5; Mismatches 4; Indels 7; Gaps 1;

Qy 5 PRLQNOQHLPALP-----IWFL 24
Db 79 PKLLQNMQSQVPSIPYAGCLSQIYFFL 105

RESULT 7
JC7708
tocopherol associated protein - human
N:Alternate names: alpha-tocopherol associated protein
C:Species: Homo sapiens (man)
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Jul-2004
C:Accession: JC7708
R:Yamauchi, J.; Iwamoto, T.; Kida, S.; Masushige, S.; Yamada, K.; Esashi, T.
Biochem. Biophys. Res. Commun. 285, 295-299, 2001
A:Title: Tocopherol-associated protein is a ligand-dependent transcriptional activator.
A:Reference number: JC7708; MUID:21338208; PMID:11444841
A:Contents: Liver
A:Accession: JC7708
A:Molecule type: mRNA
A:Residues: 1-403 <YAM>
A:Cross-references: UNIPROT:O76054; GB:AL096881
C:Comment: This protein has the following properties: (i) alpha-tocopherol specific bindi

```


A;Title: Expression of members of the putative olfactory receptor gene family in mammalian olfactory epithelium
A;Reference number: S20571; MID:92131132; PMID:1370859
A;Accession: S20571
A;Status: preliminary
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-313 <PAR>
A;Cross-references: UNIPROT:P30955; EMBL:X64996; NID:G990; PIDN:CAA46129.1; PID:9891
C;Superfamily: olfactory receptor OR14
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match 29.0%; Score 51; DB 2; Length 313;
Best Local Similarity 53.3%; Pred. No. 23;
Matches 8; Conservative 4; Mismatches 3; Indels

Qy 5 PRLLNQQQHLFALP 19
|:|:|:|:|:|:|:
Db 79 PKLLONMOSOVPSIP 93

Search completed: November 10, 2004, 13:40:45
Job time : 8.12953 secs

Query Match 29.3%; Score 51.5; DB 2; Length 240;
Best Local Similarity 48.1%; Pred. No. 15;
Matches 13; Conservative 3; Mismatches 6; Indels 5; Gaps 2;

Qy 1 RTLQPRLLNQQQHLPALPIWFLLQWL 27
Db 50 RTLIPAMLINMQ---PSLAIW--LHWI 71

RESULT 13
A37286
Olfactory receptor I15 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Jan-1993 #sequence_revision 22
C:Accession: A37286
R:Buck, L.; Axel, R.
Cell 65, 175-187, 1991
A:Title: A novel multigene family may encode odorant receptors: a molecular basis for od
A:Reference number: A23701; MUID:91191556; PMID:1840504
A:Accession: A37286
A:A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-314 #EUC>
A:A:Cross-references: UNIPROT:P23274; GB:M64392; NID:G205845; PIDN:AAA41755.1; PID:G205846
C:Superfamily: olfactory receptor OR14
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 29.3%; Score 51.5; DB 2; Length 314;
Best Local Similarity 34.4%; Pred. No. 20;
Matches 11; Conservative 7; Mismatches 7; Indels 7; Gaps 1;

QY 5 PRLNQOQHLPALP-----IWFLQLWL 29
|:||||| :|:| :|:| :|:
79 PKLNMOSOVPSIPFAGLTOLYFYLYFADL 110
Ddb

RESULT 14
E75219
Hypothetical protein PAB2167 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: E75219
R:anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: E75219
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-219 <RAW>
A:Cross-references: UNIPROT:Q9V1Z4; GB:AJ248283; GB:AL096836; NID:G5457433; PIDN:CAB49320
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2167

Query Match 29.0%; Score 51; DB 2; Length 219;
Best Local Similarity 31.8%; Pred. No. 16;
Matches 7; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

7 LLQOQQHLPALPIWFLQLWR 28
||::|:||:||||:
22 LLEHDSOIIPGIGLYVILKWAR 43

RESULT 15
S20571
olfactory receptor - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 09-Jul-2004
C:Accession: S20571
C:Parmentier, M.; Libert, F.; Schurmans, S.; Schiffmann, S.; Lefort, A.; Eggerickx, D.;
Nature 355, 453-455, 1992

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 39.0743 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-37

Perfect score: 176
Sequence: 1 RYLQRLNQOQHLPALPIWFLQWLRHPL 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	38.1	1360	2 Q9TYK4	Q9TYK4 caenorhabdi
2	62	35.2	703	2 Q7R1B4	Q7R1B4 giardia lam
3	60	34.1	225	2 Q8YTX5	Q8YTX5 anabaena sp
4	59	33.5	575	2 Q7MX22	Q7MX22 photorhabdu
5	57.5	32.7	320	2 Q34086	Q34086 coccyzus er
6	57.5	32.7	567	1 QPV_MOUSE	Q08742 mus musculu
7	57	32.4	111	2 Q89FS3	Q89FS3 bradyrizzob
8	57	32.4	352	2 Q9B214	Q9B214 calman croc
9	57	32.4	392	2 Q6PD61	Q6PD61 homo sapien
10	57	32.4	392	2 AAH58915	AAH58915 homo sapi
11	57	32.4	394	2 Q9W0M3	Q9W0M3 drosophila
12	57	32.4	401	2 Q86NX2	Q86NX2 drosophila
13	57	32.4	403	1 S142_HUMAN	Q76054 homo sapien
14	57	32.4	403	2 CAG30457	CAG30457 homo sapi
15	56.5	32.1	387	1 S142_BOVIN	P58875 bos taurus
16	56.5	32.1	400	1 S143_HUMAN	Q9UDX4 homo sapien
17	56.5	32.1	400	1 S143_RAT	Q9Z1J8 rattus norv
18	56.5	32.1	400	2 Q61S82	Q61S82 homo sapien
19	56.5	32.1	400	2 Q6XC17	Q6XC17 homo sapien
20	56.5	32.1	400	2 AAOS2677	AAOS2677 homo sapi
21	56.5	32.1	400	2 AAH69641	AAH69641 homo sapi
22	56.5	32.1	403	1 S142_RAT	Q99MS0 rattus norv
23	56.5	32.1	403	2 Q867A0	Q867A0 bos taurus
24	56.5	32.1	567	2 Q9QZU3	Q9QZU3 mus musculu
25	56	31.8	155	2 Q7S5B3	Q7S5B3 neurospora
26	56	31.8	155	2 CAF05898	CAF05898 neurospor
27	55.5	31.5	111	2 Q7QV42	Q7QV42 giardia lam
28	55	31.2	162	2 Q9KMZ3	Q9KMZ3 vibrio chol
29	55	31.2	544	2 Q8EF60	Q8EF60 shewanella
30	55	31.2	1556	2 Q60306	Q60306 homo sapien
31	54.5	31.0	406	1 S144_HUMAN	Q9UDX3 homo sapien

32	53.5	30.4	221	2 Q9QWV7	Q9QWV7 rattus norv
33	53.5	30.4	221	2 Q9QWV8	Q9QWV8 rattus norv
34	53.5	30.4	314	1 QLF9_RAT	Q23272 rattus norv
35	53.5	30.4	314	2 P70526	P70526 rattus norv
36	53.5	30.4	314	2 Q8VGI1	Q8VGI1 mus musculu
37	53	30.1	154	2 Q8E906	Q8E906 shewanella
38	53	30.1	354	2 Q8HG48	Q8HG48 alligator s
39	53	30.1	548	2 Q88J54	Q88J54 pseudomonas
40	53	30.1	549	2 P88913	P88913 human herpe
41	53	30.1	552	2 P97871	P97871 mus musculu
42	53	30.1	563	1 RK_HUMAN	Q15835 homo sapien
43	53	30.1	563	2 Q9APX3	Q9APX3 pseudomonas
44	53	30.1	563	2 AAT00534	AAT00534 homo sapi
45	53	30.1	706	2 Q7N2F5	Q7N2F5 photorhabdu

ALIGNMENTS

RESULT 1
Q9TYK4 PRELIMINARY; PRT; 1360 AA.
AC Q9TYK4; 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Hypothetical protein Y8A9A.2.
GN ORENAMES=Y8A9A.2;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851316;
RA Wilson R.;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Courtney L., Langston Y., Drone K., Mead K.;
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RT Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA EMBL, AF135463; AAK18995.1; -;
RT F1R, T33922; T3322.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA HSP; P07996; ILSL.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA WormPep; Y8A9A.2; CE21336.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA InterPro; IPR00884; TSP1.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA InterPro; IPR008085; TSP1.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Pfam; PF00090; TSP1; 5.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA PRINTS; PR01705; TSP1REPEAT.
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA PROSITE; PS00092; TSP1; 5.
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Hypothetical protein.
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Query Match 38.1%; Score 67; DB 2; Length 1360;
Best Local Similarity 47.4%; Pred. No. 4.2;
Matches 18; Conservative 3; Mismatches 8; Gaps 1;
QY 3 LQFRLNQOQHLPALPIWFL-----LQWLRHPL 32

RESULT 4	Q7MXZ2	PRELIMINARY;	PRT;	575 AA.
ID	Q7MXZ2	PRELIMINARY;	PRT;	575 AA.
AC	Q7MXZ2			
DT	01-WAR-2004	(TrEMBLrel. 26, Created)		
DT	01-WAR-2004	(TrEMBLrel. 26, last sequence update)		
DT	01-WAR-2004	(TrEMBLrel. 26, last annotation update)		
DE	Thiol:disulfide interchange protein dsbD (C-type cytochrome biogenesis protein cyzC) (Inner membrane copper tolerance protein).			
DE	Name:dsbD; OrderedLocusNames:plu4139;			
GN	Photorhabdus luminescens (subsp. laumondii).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Photorhabdus.			
OX	NCBI_TaxID=141679;			
[1]				
RN	SEQUENCE FROM N.A.			
RP	STRAIN=TT01;			
RC	MEDLINE=22957627; PubMed=14528314;			
RX	Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A., Tauric S., Bocs S., Bourreau-Eude C., Chandelier M., Charles J.-F., Dassa E., Derosé R., Derzelle S., Freysinet G., Gaudriault S., Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V., Zouine M., Glaser P., Boenare N., Danchin A., Kunst F.;			
RA	"The genome sequence of the entomopathogenic bacterium Photorhabdus luminescens."			
RT	Nat. Biotechnol. 21:1307-1313(2003).			
RL	EMBL; BX571872; CAE16511.1; -.			
DR	PhotoList; plu4139; -.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005489; F:electron transporter activity; IEA.			
DR	GO; GO:0017004; P:cytochrome biogenesis; IEA.			
DR	GO; GO:0006118; P:electron transport; IEA.			
DR	InterPro; IPR003834; Cytocb_TM.			
DR	InterPro; IPR006662; ThioRed.			
DR	InterPro; IPR006663; ThioRedox_dom2.			
DR	Pfam; PF02683; Dsbd; 1.			
DR	Pfam; PF00085; ThioRedoxin; 2.			
DR	PRINTS; PR00421; THIOREDOXIN.			
DR	PROSITE; PS00194; THIOREDOXIN; 1.			
KW	Complete proteome; Redox-active center.			
SEQUENCE	575 AA; 64072 MW; CBEC7F62BF5E7852 CRC64;			
Query Match	33.5%; Score 59; DB 2; Length 575;			
Best Local Similarity	50.0%; Pred. No. 20;			
Matches	10; Conservative	2; Mismatches	8; Indels	0; Gaps
QY	5 PRLLNQOQHLPALPTWFL 24			
DB	163 PESASNDQOHLFPSPWL 182			
	----	----	----	----
RESULT 5				
Q34086				
ID	Q34086	PRELIMINARY;	PRT;	320 AA.
AC	Q34086			
DT	01-NOV-1996	(TrEMBLrel. 01, Created)		
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)		
DT	01-WAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE	Cytochrome b (Fragment).			
DE	Coccyzus erythrophthalmus (Black-billed cuckoo).			
OC	Mitochondrion.			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Cuculiformes; Coccyzidae; Coccyzus.			
NCBI_TaxID=33604;				
[1]				
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=94356264; PubMed=8075835;			
RX	Avise J.C., Nelson W.S., Sibley C.G.;			
RA	"Why one kinase sequences from mitochondrial DNA fail to solve the			
RT	Hostin phylogenetic enigma."			
MO	Mol. Phylogenet. Evol. 3:1175-184(1994).			
CC	!- FUNCTION: Component of the ubiquinol-cytochrome c reductase			

[illegible]

```

RT Bradyrhizobium japonicum USDA110.1;
RL DNA Res. 9:189-197(2002).
DR EMBL: AP005959; BAC51891.1; -.
KW Complete proteome.
SQ SEQUENCE 111 AA; 12601 MW; 82F95EF338B88AB4 CRC64;

Query Match 32.4%; Score 57; DB 2; Length 111;
Best Local Similarity 39.3%; Pred. No. 6.5;
Matches 11; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Qy 5 PRLLNQOQHLPALPIWFLQWLRLHPL 32
Db 38 PKNTGQESWLPRILRWLRWLSRLPPL 65

RESULT 8
Q9B214 Q9B214 PRELIMINARY; PRT; 352 AA.
AC Q9B214
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE NADH dehydrogenase subunit 2.
GN Name=nadh2;
GN Name=ndh2;
OS Caiman crocodilus (Spectacled caiman) (Caiman sclerops).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Crocodylidae; Alligatorinae; Caiman.
OX NCBI_TaxID=8499;
RN NCBI_TaxID=8499;
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=21190842; PubMed=11297180;
RA Janke A., Erpenbeck D., Nilsson M., Arnason U.;
RT "The mitochondrial genomes of the iguana (Iguana iguana) and the
RT caiman (Caiman crocodilus): implications for amniote phylogeny.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 268:623-631(2001).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (By similarity).
CC -1- SIMILARITY: Belongs to the complex I subunit 2 family.
DR EMBL: AJ404872; CAC36944.1; -.
DR GO: GO:0005739; C:mitochondrion; IEA.
DR GO: GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0042773; P:ATP synthesis coupled electron transport; IEA.
DR GO: GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro: IPR003917; NADHub_oxred2.
DR InterPro: IPR010933; NADH dehy_S2_C.
DR InterPro: IPR001750; Oxidored G1.
DR Pfam: PF06444; NADH dehy_S2_C; 1.
DR Pfam: PF00361; Oxidored G1; 1.
DR PRINTS: PR01436; NADPHGNASE2.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 352 AA; 38810 MW; 68197AA6DD36D8A3 CRC64;

Query Match 32.4%; Score 57; DB 2; Length 352;
Best Local Similarity 46.4%; Pred. No. 22;
Matches 13; Conservative 3; Mismatches 8; Indels 4; Gaps 1;

Qy 5 PRLLNQOQHLPALPIWFLQWLRLHPL 32
Db 115 FEVLQ-----MPTLPSIFLLTWKLGFL 138

RESULT 9
Q6PD61 Q6PD61 PRELIMINARY; PRT; 392 AA.
AC Q6PD61
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collings P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Strausberg R.L., Collins P.S., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC058915; AAH58915.1; -.
DR InterPro: IPR001251; CRAL_TRIO_C.
DR InterPro: IPR008273; CRAL_TRIO_N.
DR InterPro: IPR001071; RetBind/tocTrans.
DR InterPro: IPR011074; Sec14p_like_N.
DR InterPro: IPR008983; TNF_like.
DR Pfam: PF00650; CRAL_TRIO; 1.
DR Pfam: PF03765; CRAL_TRIO_N; 1.
DR PRINTS: PR00180; CRETINALDHP.
DR SMART: SM00516; SEC14; 1.
DR PROSITE: PS0191; CRAL_TRIO; 1.
KW Hypothetical protein.
SQ SEQUENCE 392 AA; 44747 MW; 11A833345968FESA CRC64;

Query Match 32.4%; Score 57; DB 2; Length 392;
Best Local Similarity 44.4%; Pred. No. 25;
Matches 16; Conservative 3; Mismatches 7; Indels 10; Gaps 2;

Qy 3 LQPR-----LLQOQOQHLPALP---IFWLLQWLRL 28
Db 8 LSPRQKALAKFRENVDVLPALPNPDYFLRLRLR 43

RESULT 10
AAH58915 AAH58915 PRELIMINARY; PRT; 392 AA.
AC AAH58915
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

```

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pahay J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.T., Skalska J., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
KW EMBL; BC058915; AAHS8915.1; -
SQ SEQUENCE 392 AA; 44747 MW; 11A833345968FE5A CRC64;

Query Match 32.4%; Score 57; DB 2; Length 392;
Best Local Similarity 44.4%; Pred. No. 25;
Matches 16; Conservative 3; Mismatches 7; Indels 10; Gaps 2;

QY 3 LQPR-----LLQOQOHLPALP---TWFLQWLNR 28
DB 8 LSPROKEALAFRENVQVLPALPNPDYFLRLNR 43

RESULT 11
Q9W0M3
ID Q9W0M3 PRELIMINARY; PRT; 394 AA.
AC Q9W0M3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE CG13902-PA.
GN ORFNames=CG13902;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blasei R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Swirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Swirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Masra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003469; AAF47420.1; -
DR FlyBase; FBgn0035166; CGI3902.
DR InterPro; IPR007113; Cupin region.
DR InterPro; IPR003347; TF_JmJC.
DR Pfam; PF02373; JmJC; 1.
DR SMART; SM00558; JmJC; 1.
SQ SEQUENCE 394 AA; 44323 MW; 03E15F55F25EF503 CRC64;

Query Match 32.4%; Score 57; DB 2; Length 394;
Best Local Similarity 54.2%; Pred. No. 25;
Matches 13; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 4 QPRLNQOQHLPALPIWFLQWL 27
 DB 176 QPTLLNTTQHWPALHKWLDLNYL 199

RESULT 12
 O86NX2 PRELIMINARY; PRT; 401 AA.
 AC O86NX2;
 DT 01-JUN-2003 (TREMELrel. 24, Created)
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
 DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
 DE G21055P.
 GN ORFNames=CG13902;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Chape M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Paclet J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BT003596; AAO39599.1;
 DR FlyBase: FBgn0035166, CG13902.
 DR InterPro: IPR007113; Cupin_region.
 DR InterPro: IPR003347; TP_JmJC.
 DR Pfam: PF02373; JmJC; 1.
 DR SMART: SM00558; JmJC; 1.
 SQ SEQUENCE 401 AA; 45199 MW; 8380B46D7BF05032 CRC64;

Query Match 32.4%; Score 57; DB 2; Length 401;
 Best Local Similarity 54.2%; Pred. No. 26;
 Matches 13; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 4 QPRLNQOQHLPALPIWFLQWL 27
 DB 183 QPTLLNTTQHWPALHKWLDLNYL 206

RESULT 13
 S142 HUMAN
 ID S142 HUMAN STANDARD; PRT; 403 AA.
 AC O76054; O9ULN4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE SEC14-like protein 2 (Alpha-tocopherol associated protein) (TAP)
 DE (TAP) (supernatant protein factor) (spF) (Squalene transfer protein).
 GN Name=SEC14I2; Synonyms=KIAA1186;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORM 1), ALTERNATIVE SPLICING, AND
 RP CHARACTERIZATION.
 RX MEDLINE=20400504; PubMed=10829015; DOI=10.1074/jbc.M000851200;
 RA Zimmer S., Stocker A., Sarbelouki M.N., Spycher S.E., Sassoon J.,
 RA Azzi A.;
 RT "A novel human tocopherol-associated protein: cloning, in vitro
 RT expression, and characterization."
 RL J. Biol. Chem. 275:25672-25680(2000).
 RN [2]
 RN SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
 RC TISSUE=Liver;
 RX MEDLINE=21338208; PubMed=11444841; DOI=10.1006/bbrc.2001.5162;
 RA Yamauchi J., Iwamoto T., Kida S., Masushige S., Yamada K., Esashi T.;
 RT "Characterization of cDNA clones selected by the Genemark analysis
 RT from size-fractionated cDNA libraries from human brain."
 RL DNA Res. 6:329-336(1999).
 RL -I- FUNCTION: Carrier protein. Binds to some hydrophobic molecules and

"Tocopherol-associated protein is a ligand-dependent transcriptional
 activator."
 RL Biochem. Biophys. Res. Commun. 285:295-299(2001).
 RN [3]
 RN SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
 RP TISSUE=Liver;
 RX PubMed=11226224; DOI=10.1073/pnas.041620398;
 RA Shibata N., Arita M., Misaki Y., Dohmae N., Takio K., Ono T.,
 RA Inoue K., Arai H.;
 RT "Supernatant protein factor, which stimulates the conversion of
 RT squalene to lanosterol, is a cytosolic squalene transfer protein and
 RT enhances cholesterol biosynthesis."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:2244-2249(2001).
 RN [4]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RA Collins J.E., Huckle B.J.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208; DOI=10.1038/990031;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
 RA Clamp M., Slink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Bagghley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R., Dawson E.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson B.,
 RA Dhami P.D., Dockree C., Doddsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Levensha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Wuldar M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tilihan Y., Wright H.;
 RT "The DNA sequence of human chromosome 22."
 RL Nature 402:489-495(1999).
 RN [6]
 RN SEQUENCE OF 191-360 FROM N.A. (ISOFORM 2).
 RP TISSUE=Brain;
 RX MEDLINE=20039618; PubMed=10574461;
 RA Hirosewa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
 RA Ohara O.;
 RT "Characterization of cDNA clones selected by the Genemark analysis
 RT from size-fractionated cDNA libraries from human brain."
 RL DNA Res. 6:329-336(1999).

CC promotes their transfer between the different cellular sites.
CC Binds with high affinity to alpha-tocopherol. Also binds with a
CC weaker affinity to other tocopherols and to tocotrienols. May have
CC a transcriptional activatory activity via its association with
CC alpha-tocopherol. Probably recognizes and binds some squalene
CC structure, suggesting that it may regulate cholesterol
CC biosynthesis by increasing the transfer of squalene to a metabolic
CC active pool in the cell.
CC
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic in absence of alpha-tocopherol,
CC and nuclear in presence of alpha-tocopherol.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=O76054-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O76054-2; Sequence=VSP_006031;
CC Name=3;
CC IsoId=O76054-3; Sequence=Not described;
CC -!- TISSUE SPECIFICITY: Widely expressed. Strong expression in liver,
CC brain and prostate.
CC -!- DEVELOPMENTAL STAGE: Low expression in fetal tissues.
CC -!- SIMILARITY: Contains 1 CRAL-TRIO domain.
CC -!- SIMILARITY: Contains 1 GOLD domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AL096881; CAB51405.1; -;
CC EMBL; AC004832; AAF19256.1; -;
CC EMBL; AB033012; BAA86500.1; -;
CC FIR; JC7708; JC7708.
CC FDB; 106U; X-ray; A/C/E-1-403.
CC Genew; HGNC:10699; SEC14L2.
CC MIM; 607558; -;
CC GO; GO:0005737; Cytoplasm; NAS.
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0005386; F:carrier activity; NAS.
CC GO; GO:0005543; F:phospholipid binding; NAS.
CC GO; GO:0008431; F:tocopherol binding; NAS.
CC GO; GO:0016563; F:transcriptional activator activity; NAS.
CC GO; GO:0045893; P:positive regulation of transcription, DNA-d. . .; NAS.
CC GO; GO:0045540; P:regulation of cholesterol biosynthesis; NAS.
CC InterPro; IPR001251; CRAL-TRIO_C.
CC InterPro; IPR008273; CRAL-TRIO_N.
CC InterPro; IPR009038; GOLD.
CC InterPro; IPR01071; RetBind/tocTrans.
CC InterPro; IPR011074; Sec14p like_N.
CC InterPro; IPR008983; TNF like.
CC Pfam; PF00650; CRAL-TRIO; 1.
CC Pfam; PF03765; CRAL-TRIO_N; 1.
CC PRINTS; PR00180; CRAL-TRIO_N; 1.
CC SMART; SM00516; SEC14; 1.
CC PROSITE; PS00191; CRAL-TRIO; 1.
CC PROSITE; PS00866; GOLD; 1.
CC 3D-structure; Activator; Alternative splicing; Lipid-binding;
KW Nuclear protein; Transcription regulation; Transport.
FT DOMAIN 76 249
FT VARSPLIC 275 383
FT DOMAIN 361 403
FT VARSPLIC 361 403
FT Missing (in isoform 2).
FT Y -> H (in Ref. 2).
FT CONFLICT 36 36
FT SEQUENCE 403 AA; 46145 MW; D846747EC8D1513E CRC64;
Query Match 32.4%; Score 57; DB 1; Length 403;
Best Local Similarity 44.4%; Pred. No. 26;
Matches 16; Conservative 3; Mismatches 7; Indels 10; Gaps 2;

QY 3 LQPR-----LLNQOQHLPALP-----IWFLQWLRL 28
DB 8 LSPRQKEALAKRENVQDVLPAIPNDPDYFLLRLWL 43
RESULT 14
CAG30457 PRELIMINARY; PRT; 403 AA.
ID CAG30457
AC CAG30457;
DT 01-JUN-2004 (TrEMBLrel. 27, Created)
DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE SEC14L2 protein.
GN SEC14L2
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,
RA Cole C.G., Goward M.E., Aguado B., Mallya M., Mokrab Y., Huckle B.J.,
RA Beare D.M., Dunham I.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR456571; CAG30457.1; -;
SQ SEQUENCE 403 AA; 46145 MW; D846747EC8D1513E CRC64;
Query Match 32.4%; Score 57; DB 2; Length 403;
Best Local Similarity 44.4%; Pred. No. 26;
Matches 16; Conservative 3; Mismatches 7; Indels 10; Gaps 2;

QY 3 LQPR-----LLNQOQHLPALP-----IWFLQWLRL 28
DB 8 LSPRQKEALAKRENVQDVLPAIPNDPDYFLLRLWL 43

RESULT 15
S142 BOVIN
ID S142 BOVIN STANDARD; PRT; 387 AA.
AC P58875;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE SEC14-like protein 2 (Alpha-tocopherol associated protein) (TAP)
DE (BTAP) (Fragment).
GN Name=SEC14L2;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Meadus J., MacInnis R., Dubeski P., Hidiroglou N., Madere R.;
RT "Induction of hepatic tocopherol associated protein (TAP) mRNA but not
RT alpha-tocopherol transfer protein (TTP) mRNA in cattle fed increasing
RT levels of vitamin E.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 178-195 AND 335-353.
RP MEDLINE=20400504; PubMed=10929015; DOI=10.1074/jbc.M00851200;
RA Zimmer S., Stocker A., Sarbolouki M.N., Spycher S.E., Sassoon J.,
RA Azzi A.;
RT "A novel human tocopherol-associated protein: cloning, in vitro
RT expression, and characterization."
RL J. Biol. Chem. 275:25672-25680(2000).
CC -!- FUNCTION: Carrier protein. Binds to some hydrophobic molecules and
CC promotes their transfer between the different cellular sites.
CC Binds with high affinity to alpha-tocopherol. Also binds with a
CC weaker affinity to other tocopherols and to tocotrienols. May have
CC a transcriptional activatory activity via its association with
CC alpha-tocopherol. Probably recognizes and binds some squalene

Result No.	Score	Query Match	Length	DB	ID	Description
1	53	28.3	307	4	US-09-583-110-4473	Sequence 4473, Ap
2	53	28.3	904	4	US-09-248-796A-15547	Sequence 15547, A
3	51.5	27.5	95	4	US-09-248-796A-37928	Sequence 27928, A
4	50	26.7	631	4	US-09-830-217-12	Sequence 12, Appl
5	50	26.7	750	4	US-09-248-796A-14886	Sequence 14886, A
6	49	26.2	610	1	US-08-212-188-2	Sequence 2, Appl
7	49	26.2	610	3	US-08-970-725-2	Sequence 2, Appl
8	49	26.2	610	5	PCT-US95-02708-2	Sequence 2, Appl
9	48	25.7	160	4	US-09-270-767-38576	Sequence 38576, A
10	48	25.7	180	4	US-09-270-767-53793	Sequence 53793, A
11	48	25.7	336	4	US-09-248-796A-14171	Sequence 14171, A
12	48	25.7	414	4	US-09-252-991A-18690	Sequence 18690, A
13	47.5	25.4	151	4	US-09-270-767-33178	Sequence 33178, A
14	47.5	25.4	151	4	US-09-270-767-48395	Sequence 48395, A
15	47.5	25.4	308	4	US-09-248-796A-14928	Sequence 14928, A
16	47	25.1	1586	4	US-09-538-092-1171	Sequence 1171, Ap
17	46.5	24.9	234	4	US-09-248-796A-19500	Sequence 19500, A
18	46	24.6	110	4	US-09-247-767-61329	Sequence 61329, A
19	46	24.6	143	1	US-07-710-361-9	Sequence 9, Appl
20	46	24.6	172	4	US-09-270-767-45804	Sequence 45804, A
21	46	24.6	365	4	US-09-248-796A-18106	Sequence 18106, A
22	46	24.6	432	4	US-09-489-039A-13147	Sequence 13147, A
23	46	24.6	570	3	US-08-991-941-2	Sequence 2, Appl
24	46	24.6	570	3	US-09-173-151A-36	Sequence 36, Appl
25	46	24.6	910	4	US-09-313-943-28	Sequence 28, Appl
26	46	24.6	1079	1	US-08-455-543A-55	Sequence 55, Appl
27	46	24.6	1079	2	US-08-223-3050-55	Sequence 55, Appl

QY 12 DSCGPKLSLNKILVGNLFSSSF 35
Db 527 ETCGELLVGMGNHIRSTGNLPPDF 550

Query Match 28.3%; Score 53; DB 4; Length 904;
Best Local Similarity 41.7%; Pred. No. 16;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

US-09-248-796A-15547

RESULT 3

US-09-248-796A-27928
; Sequence 27928, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 27928
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-27928

Query Match 27.5%; Score 51.5; DB 4; Length 95;
Best Local Similarity 32.4%; Pred. No. 1.4;
Matches 12; Conservative 6; Mismatches 12; Indels 7; Gaps 1;

QY 8 LSQDSCGPKLSLNKILVGNLFSSSFT 37
Db 39 ITHKASCKDLSSSRNKLAPLIMTDVGLFLDFT 75

RESULT 4

US-09-830-217-12
; Sequence 12, Application US/09830217
; Patent No. 6521441
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Staphylococcus aureus Genes and Polypeptides
; FILE REFERENCE: PB461PCT
; CURRENT FILING DATE: 2001-04-24
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US99/06199
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/080,296
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/084,674
; PRIOR FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-830-217-12

Query Match 26.7%; Score 50; DB 4; Length 691;
Best Local Similarity 34.6%; Pred. No. 34;
Matches 9; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 1 MAVINELSQDSCGPKLSLNKIL 26
Db 551 IGLTHSTNKDVGPLKKKINGTVL 576

RESULT 5

US-09-248-796A-14886
; Sequence 14886, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14886
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14886

Query Match 26.7%; Score 50; DB 4; Length 750;
Best Local Similarity 39.0%; Pred. No. 38;
Matches 16; Conservative 3; Mismatches 18; Indels 4; Gaps 2;

QY 1 MAVIIN--ELSQDSCG--PLKISLNKILVGNLFSSFT 37
Db 420 LAEIIINSELAQSDGTGKIPESVELANKFLAVQTFSLYIMP 460

RESULT 6

US-08-212-188-2
; Sequence 2, Application US/08212188
; Patent No. 5689039
; GENERAL INFORMATION:
; APPLICANT: BECKER, JEFFREY M.
; APPLICANT: STACEY, GARY
; TITLE OF INVENTION: PLANT PEPTIDE TRANSPORT GENE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWREY & SIMON
; STREET: 1299 PENNSYLVANIA AVE., N.W.
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,188
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: AUERBACH, JEFFREY I
; REGISTRATION NUMBER: 32,680
; REFERENCE/DOCKET NUMBER: 7493-006
; TELEPHONE: (202) 383-7451
; TELEFAX: (202) 383-6610
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

Db 3 VSMADNCGNURLSNLNSGNEYLPSANAYGNQYEDF 38

RESULT 10

US-09-270-767-53793

; Sequence 53793, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 53793

; LENGTH: 160

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-270-767-53793

Query Match 25.7%; Score 48; DB 4; Length 160;

Best Local Similarity 33.3%; Pred.No.10;

Matches 12; Conservative 7; Mismatches 9; Indels 8; Gaps 2;

QY 8 LSORDSCGPLKISLN---NKIL---VYGNLFSSF 35

Db 3 VSMADNCGNURLSNLNSGNEYLPSANAYGNQYEDF 38

RESULT 11

US-09-248-796A-14171

; Sequence 14171, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 14171

; LENGTH: 336

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-14171

Query Match 25.7%; Score 48; DB 4; Length 336;

Best Local Similarity 33.3%; Pred.No.27;

Matches 11; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 5 INELSRDSCGPLKISLNKILVYGNLFSSFTF 37

Db 272 LGNVSRDEKPKDGISFENVKLAFAADALSRYTP 304

RESULT 12

US-09-252-991A-18690

; Sequence 18690, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

```

||||| |||| |::: :||
71 RDSQCQSGGPLVTSIDGQLKLYG 94

RESULT 15
US-09-248-796A-14928
; Sequence 14928, Application US/09248796A
; Patent No 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14928
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14928

Query Match      25.4%; Score 47.5; DB 4; Length 308;
Best Local Similarity 44.4%; Pred. No. 29;
Matches 12; Conservative 4; Mismatches 7; Gaps 1;

QY      15 GPLK-----ISLNKKILVYGNIFFS 34
          ||::: ||::: ||::: ||::: ||:::
Db       101 GPVENIAISSWIKINDRIITLTKLFSS 127

Search completed: November 10, 2004, 13:44:09
Job time : 13.7807 secs
```

This Page Blank (uspto)

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 39.7478 Seconds
(without alignments)
328.807 Million cell updates/sec

US-10-092-750-38

Perfect score: 187

Sequence: 1 MAVIINELSQDSCGPKISLNKILVYGNLFSSFTP 37

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 segs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubaa/US10D_PUBCOMB.pep.*

17: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*

18: /cgn2_6/ptodata/1/pubaa/US11_NEW_PUB.pep.*

19: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*

20: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	187	100.0	37	14	US-10-092-750-38
2	57	30.5	1307	15	US-10-282-122A-78432
3	55	29.4	619	14	US-10-369-493-5855
4	54	28.9	324	14	US-10-369-493-10198
5	52	27.8	755	9	US-09-738-626-4773
6	51.5	27.5	122	17	US-10-425-115-311744
7	51.5	27.5	158	17	US-10-425-115-267293
8	51.5	27.5	358	15	US-10-425-114-65925
9	51.5	27.5	437	17	US-10-425-115-311741
10	51.5	27.5	441	15	US-10-425-114-64080
11	51	27.3	96	17	US-10-425-115-198064
12	50.5	27.0	66	15	US-10-424-599-280511
13	50	26.7	183	16	US-10-437-963-126350

ALIGNMENTS

RESULT 1

US-10-092-750-38

; Sequence 38, Application US/10092750

; Publication No. US20030032157A1

; GENERAL INFORMATION:

; APPLICANT: Hammond, Philip W.

; APPLICANT: Alpin, Julia

; APPLICANT: Wright, Martin C.

; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1

; FILE REFERENCE: 50036/050002

; CURRENT APPLICATION NUMBER: US/10/092,750

; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 60/274,526

; PRIOR FILING DATE: 2001-03-08

; NUMBER OF SEQ ID NOS: 253

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 38

; LENGTH: 37

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-092-750-38

Query Match 100.0%; Score 187; DB 14; Length 37;

Best Local Similarity 100.0%; Pred. No. 3.6e-21;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVIINELSQDSCGPKISLNKILVYGNLFSSFTP 37

DB 1 MAVIINELSQDSCGPKISLNKILVYGNLFSSFTP 37

RESULT 2

US-10-282-122A-78432

; Sequence 78432, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

```

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78432
; TYPE: PRT
; LENGTH: 1307
; ORGANISM: Yersinia pestis
US-10-282-122A-78432

Query Match 30.5%; Score 57; DB 15; Length 1307;
Best Local Similarity 50.0%; Pred. No. 30;
Matches 13; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 7 ELQRDSCGPKLSLNKILVGNLF 32
||| : : : : : : : : : : : :
Db 416 ELGSGARLGRNINLNKSLFPYGLF 441

RESULT 3
US-10-369-493-5855
; Sequence 5855, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5855
; TYPE: PRT
; LENGTH: 1307
; ORGANISM: Yersinia pestis
US-10-282-122A-78432

```

```

; LENGTH: 619
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5855

Query Match 29.4%; Score 55; DB 14; Length 619;
Best Local Similarity 43.5%; Pred. No. 24;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 15 GPLKISLNKILVGNLFSSFTP 37
||| : : : : : : : : : : : :
Db 238 GEYSVSDGKSLYGNMDSVTP 260

RESULT 4
US-10-369-493-10198
; Sequence 10198, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10198
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Cytophaga hutchinsonii
US-10-369-493-10198

Query Match 28.9%; Score 54; DB 14; Length 324;
Best Local Similarity 41.2%; Pred. No. 15;
Matches 14; Conservative 6; Mismatches 12; Indels 2; Gaps 1;

QY 4 IINELQRDSCGPKLSLNKILVGNLFSSFTP 37
||| : : : : : : : : : : : :
Db 260 IINATAERNVGPVKVWIN--IERYGNTTSATIP 291

RESULT 5
US-09-738-626-4773
; Sequence 4773, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059

```

Matches	15;	Conservative	4;	Mismatches	11;	Indels	7;	Gaps	1;
Qy	1	MAVINELSQDSCGPKLISLNKILVYGNLFSSFT 37							
Db	28	LVGTTSLAQSSSLGRLRISPNK-----STSFT 57							
<p>RESULT 8</p> <p>US-10-425-114-65925</p> <p>Sequence 65925, Application US/10425114</p> <p>Publication No. US20040034888A1</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Liu, Jingdong</p> <p>APPLICANT: Zhou, Yihua</p> <p>APPLICANT: Kovalic, David K.</p> <p>APPLICANT: Screen, Steven E</p> <p>APPLICANT: Tabaska, Jack E</p> <p>APPLICANT: Cao, Yongwei</p> <p>TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With</p> <p>TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement</p> <p>FILE REFERENCE: 38-21(53313)B</p> <p>CURRENT APPLICATION NUMBER: US/10/425,114</p> <p>CURRENT FILING DATE: 2003-04-28</p> <p>NUMBER OF SEQ ID NOS: 73128</p> <p>SEQ ID NO 65925</p> <p>LENGTH: 358</p> <p>TYPE: PRT</p> <p>ORGANISM: Zea mays</p> <p>FEATURE:</p> <p>OTHER INFORMATION: Clone ID: LIB3606-010-G12_FLI.pep</p> <p>US-10-425-114-65925</p> <p>Query Match 27.5%; Score 51.5; DB 15; Length 358;</p> <p>Best Local Similarity 34.4%; Pred. No. 42;</p> <p>Matches 11; Conservative 7; Mismatches 7; Indels 7; Gaps 1;</p>									
Qy	6	NELSQDSCGPKLISLNKILVYGNLFSSFT 37							
Db	304	NKLADQSEC-----NNRVIDGNLITSQGP 328							
<p>RESULT 9</p> <p>US-10-425-115-311741</p> <p>Sequence 311741, Application US/10425115</p> <p>Publication No. US20040214272A1</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: La Rosa, Thomas J.</p> <p>APPLICANT: Kovalic, David K.</p> <p>APPLICANT: Zhou, Yihua</p> <p>APPLICANT: Cao, Yongwei</p> <p>TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With</p> <p>TITLE OF INVENTION: Plants</p> <p>FILE REFERENCE: 38-21(53222)B</p> <p>CURRENT APPLICATION NUMBER: US/10/425,115</p> <p>CURRENT FILING DATE: 2003-04-28</p> <p>NUMBER OF SEQ ID NOS: 369326</p> <p>SEQ ID NO 311741</p> <p>LENGTH: 437</p> <p>TYPE: PRT</p> <p>ORGANISM: Zea mays</p> <p>FEATURE:</p> <p>OTHER INFORMATION: Clone ID: MRT4577_47360C.1.pep</p> <p>US-10-425-115-311741</p> <p>Query Match 27.5%; Score 51.5; DB 17; Length 437;</p> <p>Best Local Similarity 34.4%; Pred. No. 53;</p> <p>Matches 11; Conservative 7; Mismatches 7; Indels 7; Gaps 1;</p>									
Qy	6	NELSQDSCGPKLISLNKILVYGNLFSSFT 37							
Db	383	NKLADQSEC-----NNRVIDGNLITSQGP 407							

Search completed: November 11, 2004, 01:28:20
Job time : 40.7978 secs

US-10-282-122A-53155

This Page Blank (uspio)

Db 590 AKVVEDKENPDRVG---IHLNNTSVNTGNLFRRYTP 622

RESULT 8

D81215
TonB-dependent receptor, probable NMB0293 [imported] - Neisseria meningitidis (strain MC58)
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: D81215
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vignoli, P.
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: D81215
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-703 <TET>
A:Cross-references: UNIPROT:Q9URCS; GB:AB002386; GB:AB002098; NID:97225512; PIDN:AAF4074
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0293

Query Match 28.6%; Score 53.5; DB 2; Length 703;

Best Local Similarity 38.9%; Pred. No. 17;

Matches 14; Conservative 4; Mismatches 15; Indels 3; Gaps 1;

Qy 2 AVINELSQDSCGPKLISLNKKILVYGNLFSSSTP 37

Db 590 AKVVEDKENPDRVG---IHLNNTSVNTGNLFRRYTP 622

RESULT 9

B98021
Conserved hypothetical protein sprl195 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: B98021
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Beyer, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Meyer, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21423245; PMID:11544234
A:Accession: B98021
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-307 <KUR>
A:Cross-references: UNIPROT:Q8DPF1; GB:AE007317; PIDN:AAK99998.1; PID:gl5458528; GSPDB:G09091
C:Genetics:
A:Gene: sprl195

Query Match 28.3%; Score 53; DB 2; Length 307;

Best Local Similarity 37.9%; Pred. No. 8.4; Length 307;

Matches 11; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 4 IINELSQDSCGPKLISLNKKILVYGNLF 32

Db 5 VINELDSPGLAGMCRINNRLVYGGSY 33

RESULT 10

G90091
Hypothetical protein orf670 [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: G90091
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Reid
Nature 410, 1091-1096, 2001

A:Title: The highly reduced genome of an enslaved algal nucleus.

A:Reference number: A99082; MUID:11323671; PMID:11323671

A:Accession: G90091

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-670 <DOU>

A:Cross-references: UNIPROT:Q98RT4; GB:AF165818; NID:gl13794489; PIDN:AAK39864.1; GSPDB:G

C:Genetics:

A:Gene: orf670

A:Map position: 1

A:Genome: nucleomorph

C:Keywords: nucleomorph

Query Match 28.3%; Score 53; DB 2; Length 670;

Best Local Similarity 36.6%; Pred. No. 20;

Matches 15; Conservative 7; Mismatches 9; Indels 10; Gaps 2;

Qy 5 INELSQDSCGPKLISLNKKILVYGNLFSSSF 35

Db 226 LMQNFNDIKILYHDSLNKISLGNKILVFGILLPELFDY 266

RESULT 11

E90284

Hypothetical protein SSO1294 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C:Accession: E90284

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-Y

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.

arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: E90284

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-247 <KUR>

A:Cross-references: UNIPROT:Q97YM8; GB:AE006641; NID:gl3814496; PIDN:AAK41532.1; GSPDB:G

C:Genetics:

A:Gene: SSO1294

Query Match 27.8%; Score 52; DB 2; Length 247;

Best Local Similarity 35.7%; Pred. No. 9.2;

Matches 10; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

Qy 4 IINELSQDSCGPKLISLNKKILVYGNL 31

Db 200 IVREISEVVKCDILTIPNEKLVQYANM 227

RESULT 12

T51044

related to spore coat protein SP96 precursor [imported] - Neurospora crassa

N:Alternate names: protein B15120.140

C:Species: Neurospora crassa

C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004

C:Accession: T51044

R:Schulte, J.; Aign, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, July 2000

A:Reference number: Z25286

A:Accession: T51044

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-461 <SCH>

A:Cross-references: UNIPROT:Q9P3J7; EMBL:AL389900; GSPDB:GN00116; NCSP:B15120.140

A:Experimental source: BAC clone B15120; strain OR74A

C:Genetics:

A:Gene: NCSP:B15120.140

A:Map position: 6

Query Match 27.0%; Score 50.5; DB 2; Length 461;

Best Local Similarity 37.5%; Pred. No. 30;

Result No.	Score	Query Match	Length	DB	ID	Description
1	59	31.6	3085	2	Q70710	infectious
2	57.5	30.7	1077	2	Q9m877	arabidopsals
3	57	30.5	660	2	Q8K14	dictyostali
4	57	30.5	1063	2	Q8dlP4	yersinia pe
5	57	30.5	1307	2	Q8a2W2	Q8a2W2 yersinia pe
6	57	30.5	1307	2	AAS64021	AAS64021 yersinia
7	56.5	30.2	2153	1	POLG_HRV16	Q8a122 h genome po
8	55	29.4	619	1	CALX_CAEEL	P34652 caenorhabdi
9	55	29.4	1914	2	Q6CDD3	Q6cdd3 yarrowia li
10	54.5	29.1	438	2	Q84Kv6	Q84kv6 brassica ol
11	54.5	29.1	1128	2	Q98QZ4	Q98qz4 mycoplasma
12	54	28.9	412	2	Q83V23	Q83v23 pseudomonas
13	54	28.9	504	2	Q9ZW78	Q9zw78 arabidopsals
14	54	28.9	895	2	Q94A14	Q94a14 arabidopsals
15	53.5	28.6	137	2	P95944	P95944 sulfolobus
16	53.5	28.6	702	2	Q7JPK2	Q7jpk2 neisseria m
17	53.5	28.6	703	2	Q7AX11	Q7ax11 neisseria m
18	53.5	28.6	703	2	Q9JPK1	Q9jpk1 neisseria m
19	53.5	28.6	703	2	Q9JPK3	Q9jpk3 neisseria m
20	53.5	28.6	703	2	Q9JRC5	Q9jrc5 neisseria m
21	53.5	28.6	703	2	Q7JRC7	Q7jrc7 neisseria m
22	53.5	28.6	703	2	Q7JDDR7	Q7jddr7 neisseria m
23	53.5	28.6	911	2	Q6CPS0	Q6cps0 kluyveromyc
24	53	28.3	220	2	Q9PYZ6	Q9pyz6 xestia c-ni
25	53	28.3	307	2	Q8DPF1	Q8dpf1 streptococc
26	53	28.3	330	1	OR42_MOUSE	Q8vf12 mus musculu
27	53	28.3	352	2	Q9P930	Q9p930 candida alb
28	53	28.3	670	2	Q98RT4	Q98rt4 guillardia
29	53	28.3	1201	2	Q7RHG8	Q7rhg8 plasmodium
30	52	27.8	247	2	Q97YM8	Q97ym8 sulfolobus
31	52	27.8	479	2	Q7TSF9	Q7tsf9 mus musculu

Complete proteome.
KW SEQUENCE 1307 AA; 144414 MW; F2C5A282B066CE54 CRC64;
SQ

Query Match 30.5%; Score 57; DB 2; Length 1307;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 13; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 7 ELSQRDSCGPLKISLNKILVYGNLF 32
||| ||| : ||| : |||
Db 416 ELSGSRALGRNLNINLKNSLLPYGDLF 441

RESULT 6

```

RESULT 6
AAS64021
ID AAS64021 PRELIMINARY; PRT; 1307 AA.
AC AAS64021;
DT 24-MAR-2004 (TtEMBLrel. 27, Created)
DT 24-MAR-2004 (TtEMBLrel. 27, Last sequence update)
DT 04-MAY-2004 (TtEMBLrel. 27, Last annotation update)
DE Putative exported protein.
GN yf3876.
OS yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
CX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91001 / Biovar Mediaevalis;
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,
RA Yan Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,
RA He C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,
RA Yang R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AE017142; AAS64021.1; -
SO SEQUENCE 1307 AA; 144414 MW; F2C3A282B0656C854 CR664

```

Query Match 30.5%; Score 57; DB 2; Length 1307;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 13; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 7 ELSQRDSCGPLKISLNNKILVYGNLF 32
 ||| | | | | : | | : ||
 db 416 ELSGSARLGRNLNLIKSLLPYGDLE 441

RESULT 7

RESULT 7

POLG HRV16 STANDARD; PRT; 2153 AA.

1 ID_POLG_HRV16 Q82152;

2 AT 15-JUL-1998 (Rel. 36, Created)

3 DT 15-DEC-1998 (Rel. 37, Last sequence update)

4 DT 01-OCT-2004 (Rel. 45, Last annotation update)

5 DE Genome polyprotein [Contains: Coat protein VP4 (PI4); Coat protein VP2 (PI2); Coat protein VP3 (PI3); Coat protein VP1 (PI1); Core protein P1B); Core protein P2B; Core protein P2C; Core protein P3A; Genome- P2A; Core protein VP6 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) DE 2 linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) DE 3 (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].

6 OS Human rhinovirus 16 (HRV-16).

7 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; OC Rhinovirus.

8 NCBI_TaxID=31708;

9 [1]

10 SEQUENCE FROM N.A.

11 RP MEDLINE=95250310; PubMed=7732663;

12 RA Lee W.M., Wang W., Rueckert R.R.;

13 RT "Complete sequence of the RNA genome of human rhinovirus 16, a RT clinically useful common cold virus belonging to the ICAM-1 receptor RT group.;"

14 RL Virus Genes 9:177-181(1995).

15 [2]

16 RP X-RAY CRYSTALLOGRAPHY (3.5 ÅNGSTRÖMS) OF 2-853.

17 RX MEDLINE=94345864; PubMed=7915182;

18 RA Oliveira M.A., Zhao R., Lee W.M., Kremer M.J., Minor I.,

19 RA Rueckert R.R., Diana G.D., Pevear D.C., Dutko F.J., McKinlay M.A.,

20 RA Rossmann M.G.;

21 RT "The structure of human rhinovirus 16.;"

Structure 1:51-68(1993).

[3]

X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 2-853, AND REVISIONS TO
547-548.

MEDLINE=97238938; PubMed=9083115;
Hadfield A.T., Lee W.M., Zhao R., Oliveira M.A., Minor I.,
Rueckert R.T., Rosemann M.G.,
"The refined structure of human rhinovirus 16 at 2.15-A resolution:
RT implications for the viral life cycle.";
Structure 5:427-441(1997).

CC -I- FUNCTION: Picornain 3C is a thiol protease that cleaves at certain
CC Gln-|-Gly sites in the polyprotein.
CC -I- CATALYTIC ACTIVITY: Selective cleavage of Gln-|-Gly bond in the
CC poliovirus polyprotein. In other picornavirus reactions Glu may be
CC substituted for Gln, and Ser or Thr for Gly.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).

CC -I- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC each of which is composed of one copy each of proteins VP1, VP2,
CC VP3, and VP4.
CC -I- PWM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -I- SIMILARITY: The protease belongs to peptidase family C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce>
CC or send an email to license@isb-sib.ch).

DR	EMBL; L24917; AAA69862.1; -	2=70-330, 3=331-568.
DR	PDB; 1AYM; X-ray; 1=569-853	2=70-330, 3=331-568.
DR	PDB; 1AYN; X-ray; 1=569-853	2=70-330, 3=331-568.

Best Local Similarity 50.0%; Pred.No. 62;
Matches 13; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

DR	InterPro: IPR000199; Pept_C3_picorn.		
DR	InterPro: IPR004004; Pept_Calici.		
DR	InterPro: IPR003003; Pept_Ser_Cys.		
DR	InterPro: IPR003138; Pico_P1A.		
DR	InterPro: IPR002527; Pico_P2B.		
DR	InterPro: IPR001676; Rhv.		
DR	InterPro: IPR000605; RNA_helicase.		
DR	InterPro: IPR007095; RNA_pol_DS_PS.		
DR	InterPro: IPR001205; RNA_pol_P3D.		
DR	InterPro: IPR007094; RNA_pol_PSwir.		
DR	InterPro: IPR008975; Viral_cap_coat.		
DR	Pfam: PF00548; Cys_protease_3C; 1.		
DR	Pfam: PF02226; Pico_P1A; 1.		
DR	Pfam: PF00947; Pico_P2A; 1.		
DR	Pfam: PF01552; Pico_P2B; 1.		
DR	Pfam: PF000773; Rhv; 3.		
DR	Pfam: PF00680; RNA_dep_RNA_pol; 1.		
DR	Pfam: PF00910; RNA_helicase; 1.		
DR	PRINTS: PR00918; CALICIVIRUSNS.		
DR	ProDom: PD001125; Cys_protease_3C; 1.		
DR	ProDom: PD001306; Pico_P2A; 1.		
DR	ProDom: PD001274; Pico_P2B; 1.		
DR	SMART; SM00382; AAA; 1.		
KW	3D-structure; Coat protein; Core protein; Hydrolase; Lipoprotein;		
KW	Myristate; Polyprotein; RNA-directed RNA polymerase; Thiol protease;		
KW	Transferase.	2	69
FT	CHAIN	70	Coat protein VP4.
FT	CHAIN	330	Coat protein VP2.
FT	CHAIN	331	Coat protein VP3.
FT	CHAIN	569	Coat protein VP1.
FT	CHAIN	854	Core protein p2A.
FT	CHAIN	996	Core protein p2B.
FT	CHAIN	1091	Core protein p2C.
FT	CHAIN	1413	Core protein p3A.
FT	CHAIN	1490	Core protein p3B.
FT	CHAIN	1510	Genome-linked protein VPG.
FT	CHAIN	1511	Picornain 3C.

RNA-directed RNA polymerase.
N-myristoyl glycine (by host).
Protease (Potential).
Protease (Potential).
KD -> NH (in Ref. 1).

1694 2153
FT CHAIN 2
FT LIPID 2
FT ACT_SITE 1657
FT ACT_SITE 1657
FT CONFLICT 1671
FT STRAND 547
FT STRAND 27
FT STRAND 30
FT STRAND 36
FT STRAND 81
FT STRAND 83
FT STRAND 87
FT STRAND 88
FT STRAND 90
FT STRAND 94
FT STRAND 101
FT STRAND 102
FT STRAND 103
FT STRAND 105
FT STRAND 113
FT STRAND 115
FT STRAND 123
FT STRAND 128
FT STRAND 129
FT STRAND 133
FT STRAND 134
FT STRAND 138
FT STRAND 140
FT STRAND 142
FT STRAND 143
FT STRAND 147
FT STRAND 151
FT STRAND 152
FT STRAND 153
FT STRAND 155
FT STRAND 156
FT STRAND 157
FT STRAND 159
FT STRAND 167
FT STRAND 168
FT STRAND 180
FT STRAND 186
FT STRAND 187
FT STRAND 188
FT STRAND 197
FT STRAND 198
FT STRAND 199
FT STRAND 204
FT STRAND 207
FT STRAND 213
FT STRAND 216
FT STRAND 219
FT STRAND 221
FT STRAND 223
FT STRAND 224
FT STRAND 225
FT STRAND 227
FT STRAND 232
FT STRAND 233
FT STRAND 241
FT STRAND 242
FT STRAND 245
FT STRAND 246
FT STRAND 248
FT STRAND 253
FT STRAND 256
FT STRAND 260
FT STRAND 261
FT STRAND 263
FT STRAND 266
FT STRAND 271
FT STRAND 280
FT STRAND 282
FT STRAND 284
FT STRAND 285
FT STRAND 288
FT STRAND 299
FT STRAND 303
FT STRAND 304
FT STRAND 307
FT STRAND 323
FT STRAND 341
FT STRAND 345
FT STRAND 353
FT STRAND 355
FT STRAND 356
FT STRAND 357
FT STRAND 369
FT STRAND 370
FT STRAND 372
FT STRAND 373
FT STRAND 374
FT STRAND 377
FT STRAND 378
FT STRAND 381
FT STRAND 382
FT STRAND 385
FT STRAND 386
FT STRAND 388
FT STRAND 389
FT STRAND 393
FT STRAND 395
FT STRAND 398
FT STRAND 399
FT STRAND 402
FT STRAND 408
FT STRAND 411
FT STRAND 416
FT STRAND 420
FT STRAND 422
FT STRAND 426
FT STRAND 428
FT STRAND 433
FT STRAND 434
FT STRAND 435
FT STRAND 436
FT STRAND 440
FT STRAND 443
FT STRAND 449
FT STRAND 453

FT STRAND 456
FT STRAND 458
FT STRAND 464
FT STRAND 474
FT STRAND 478
FT STRAND 479
FT STRAND 481
FT STRAND 486
FT STRAND 492
FT STRAND 497
FT STRAND 506
FT STRAND 507
FT STRAND 513
FT STRAND 514
FT STRAND 518
FT STRAND 523
FT STRAND 528
FT STRAND 531
FT STRAND 532
FT STRAND 537
FT STRAND 545
FT STRAND 548
FT STRAND 550
FT STRAND 554
FT STRAND 567
FT STRAND 567
FT STRAND 570
FT STRAND 579
FT STRAND 580
FT STRAND 584
FT STRAND 588
FT STRAND 600
FT STRAND 601
FT STRAND 602
FT STRAND 603
FT STRAND 605
FT STRAND 607
FT STRAND 615
FT STRAND 618
FT STRAND 624
FT STRAND 629

Query Match 30.2%; Score 56.5; DB 1; Length 2153;
Best Local Similarity 41.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 8; Mismatches 6; Indels 9; Gaps 3;

QY 6 NELSD--SCGLKI---SLNNKIL---VYGNLFSSP 35
DB 1853 DELRKDKIAGKRVIEASSINDTILFTVYGNLFSSP 1891

RESULT 8

CALX_CABEL STANDARD; PRT; 619 AA.
AC P34652;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Calnexin homolog precursor.
GN ORFNames-ZK632.6;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins J., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL Nature 368:32-38(1994).
CC -!- FUNCTION: Calcium-binding protein that interacts with newly
CC synthesized glycoproteins in the endoplasmic reticulum. It may act
CC in assisting protein assembly and/or in the retention within the
CC ER of unassembled protein subunits. It seems to play a major role
CC in the quality control apparatus of the ER by the retention of
CC incorrectly folded proteins (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Endoplasmic

```
CC      reticulum (By similarity).
CC      -!- SIMILARITY: Belongs to the calreticulin family.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      -----
DR      EMBL; Z22181; CAA80183.1; -.
DR      PIR; S40938; S40938.
DR      HSSP; P18418; 1K9C.
DR      IntAct; P34652; -.
DR      WormPep; ZK632.6; CE00423.
DR      InterPro; IPR001580; Calret/calnex.
DR      InterPro; IPR003033; Calret_calnex_P.
DR      InterPro; IPR008985; Coxa_like_lec_gl.
DR      Pfam; PF00262; Calreticulin; 1.
DR      PRINTS; PR00626; CALRETICULIN.
DR      ProDom; PD001866; Calreticulin; 1.
DR      PROSITE; PS00803; CALRETICULIN 1; 1.
DR      PROSITE; PS00804; CALRETICULIN 2; 1.
DR      PROSITE; PS00805; CALRETICULIN REPEAT; 1.
KW      Calcium-binding; Chaperone; Endoplasmic reticulum;
KW      Hypothetical protein; Lectin; Repeat; Signal; Transmembrane.
FT      SIGNAL      1      21      Potential.
FT      CHAIN        22      619      Calnexin homolog.
FT      CARBOHYD     203      203      N-linked (GlcNAc...) (Potential).
FT      CARBOHYD     571      571      N-linked (GlcNAc...) (Potential).
SQ      SEQUENCE      619 AA; 69207 MW; 08903CE519A75B98 CRC64;
Query Match      29.4%; Score 55; DB 1; Length 619;
Best Local Similarity      43.5%; Pred. No. 55;
Matches      10; Conservative      4; Mismatches      9; Indels      0; Gaps      0;
QY      15 GPLKISLNKKLVGNLFSSFTP 37
Db      238 GEYSVSDGKSLYYGNMMSDVT 260
RESULT 9
Q6CDD3 ID Q6CDD3 PRELIMINARY; PRT; 1914 AA.
AC Q6CDD3;
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to t[Q9UVK9 Yarrowia lipolytica Fkslp (Fragment).
GN ORFNames=VAL10C01411g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG GENOLEVURES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talia E.,
RA Goffard N., Franguel L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boursane A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz J.M., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekala F., Weslowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
```

```
RT      "Genome evolution in yeasts.";
RL      Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; CR382129; CAG81624.1; -.
FT NON TER      1
SQ SEQUENCE      1914 AA; 220365 MW; 1DE84B605332F9ED CRC64;
Query Match      29.4%; Score 55; DB 2; Length 1914;
Best Local Similarity      46.9%; Pred. No. 1.8e+02;
Matches      15; Conservative      5; Mismatches      10; Indels      2; Gaps      1;
QY      5 INELSQDSGGLKISLNKKIL--VYGNLFSS 34
Db      1596 LNEISETDSRGVMKPSLVNVLSEVVGLLSA 1627
RESULT 10
Q84KV6 ID Q84KV6 PRELIMINARY; PRT; 438 AA.
AC Q84KV6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE S-locus receptor kinase (Fragment).
GN Name=SRK-64;
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid1; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22286392; PubMed=12399400;
RA Sato K., Nishio T., Kimura R., Kusaba M., Suzuki T., Hatakeyama K.,
RA Cockendon D.J., Satta Y.;
RT "Coevolution of the S-locus genes SRK, SLG and SP11/SCR in Brassica
RL oleracea and B. rapa.";
RL Genetics 162:931-940(2002).
DR EMBL; AB054725; BAC24059.1; -.
DR GO; GO:0016301; P-kinase activity; IEA.
DR GO; GO:0004872; F-receptor activity; IEA.
DR GO; GO:0005529; F-sugar binding; IEA.
DR InterPro; IPR001480; B_lectin.
DR InterPro; IPR003609; Pan_app.
DR InterPro; IPR000858; Slocus_glycop.
DR Pfam; PF01453; B_lectin; 1.
DR Pfam; PF00354; S-locus glycop; 1.
DR SMART; SM00108; B_lectin; 1.
DR SMART; SM00473; PAN_AP; 1.
DR PROSITE; PS0927; BULB_LECTIN; 1.
DR PROSITE; PS0948; PAN; 1.
KW Kinase; Receptor.
FT NON TER      1
FT NON TER      438
FT NON TER      438
SQ SEQUENCE      438 AA; 49636 MW; 93452E2A7BE2008E CRC64;
Query Match      29.1%; Score 54.5; DB 2; Length 438;
Best Local Similarity      40.6%; Pred. No. 45;
Matches      13; Conservative      8; Mismatches      8; Indels      3; Gaps      1;
QY      1 MAVINELSQDSGGLKISLNKKILVYGNLF 32
Db      17 LSIYINTLSRES---LKISSNRTLVSPGSIF 45
RESULT 11
Q98Q24 ID Q98Q24 PRELIMINARY; PRT; 1128 AA.
AC Q98Q24;
```

DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein MYPU_2160.
 GN OrderedLocusNames-MYPU_2160;
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=1267165; PubMed=11353084;
 RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
 RA Mosser I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis.";
 RL Nucleic Acids Res. 29:2145-2153(2001).
 DR EMBL: AL445563; CAC13389.1; -;
 DR FIR; H90538; H90538.
 DR Mypulist; MYPU_2160; -;
 DR GO; GO:0004435; P:phosphoinositide phospholipase C activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006629; P:lipid metabolism; IEA.
 DR InterPro: IPR007326; Lipoprotein_17.
 DR InterPro: IPR001711; PI_PLC_Y.
 DR Pfam; PF04200; Lipoprotein_17; 4.
 DR PROSITE; PS00068; PIPLC_Y_DOMAIN; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 1128 AA; 127054 MW; 56D54050D0742EDE CRC64;

Query Match 29.1%; Score 54.5; DB 2; Length 1128;

Best Local Similarity 47.1%; Pred. No. 1.2e+02;

Matches 16; Conservative 4; Mismatches 9; Indels 5; Gaps 3;

QY 4 IINLSQRDSCGPLKISLNKILVGNLFSSFTP 37

DB 503 VANNISQRN-SSLRISFNNS---YANR-SSFTP 531

RESULT 12

ID Q83V23 PRELIMINARY; PRT; 412 AA.
 AC Q83V23;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Porin protein.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=294;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2507682; PubMed=12620844;
 RA Muraki T., Taki M., Hasegawa Y., Iwaki H., Lau P.C.K.;
 RT "Prokaryotic homologs of the eukaryotic 3-hydroxyanthranilate 3,4-
 RT dioxygenase and 2-amino-3-carboxymuconate-6-semialdehyde decarboxylase
 RT in the 2-nitrobenzoate degradation pathway of Pseudomonas fluorescens
 RT strain KU-7.";
 RL Appl. Environ. Microbiol. 69:1564-1572(2003).
 DR EMBL; AB088043; BAC534.1; -;
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005741; C:mitochondrial outer membrane; IEA.
 DR GO; GO:0015288; P:porin activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro: IPR005318; OprD.
 DR Pfam; PF03573; OprD; 1.
 KW Porin.
 SQ SEQUENCE 412 AA; 45856 MW; ACD0DA8697A1CA8B CRC64;
 Query Match 28.9%; Score 54; DB 2; Length 412;
 Best Local Similarity 50.0%; Pred. No. 49;

Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 QY 15 GPLKISLNKILVGNLFSS 34
 DB 124 GALKVAFPSNVLKYGQFMS 143

RESULT 13

ID Q9ZW78 PRELIMINARY; PRT; 504 AA.
 AC Q9ZW78;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein At2g43170.
 GN Name=At2g43170;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Renning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004450; AAC64306.1; -;
 DR FIR; H84862; H84862.
 KW Hypothetical protein.
 SQ SEQUENCE 504 AA; 52252 MW; 8B5E7FD3D6F100FB CRC64;

Query Match 28.9%; Score 54; DB 2; Length 504;

Best Local Similarity 39.3%; Pred. No. 61;

Matches 11; Conservative 6; Mismatches 9; Indels 2; Gaps 1;

QY 10 QRDSQPLKISLNKILVGNLFSSFTP 37

DB 266 QRTQSGPVTLOGNNVNM--GDMFSQATP 291

RESULT 14

ID Q94A14 PRELIMINARY; PRT; 895 AA.
 AC Q94A14;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE At2g43170/F1432.11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Koesema E., Chen H., Cheuk R., Kim C.J., Meyers M.C., Shinn P.,
 RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
 RA Chan M.N., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,

Search completed: November 10, 2004, 13:38:41
Job time : 47.1796 secs

This Page Blank (uspro)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 5 52677 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-39
Perfect score: 84
Sequence: 1 GLAKSKRNPNALTPP 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	43	51.2	461 4	US-09-248-796A-19247 Sequence 19247, A
2	42	50.0	274 4	US-09-790-045-14 Sequence 14, Appl
3	42	50.0	274 4	US-10-222-577-14 Sequence 14, Appl
4	42	50.0	274 4	US-10-222-578-14 Sequence 14, Appl
5	42	50.0	391 4	US-09-252-991A-20111 Sequence 20111, A
6	42	50.0	427 4	US-09-270-767-45426 Sequence 45426, A
7	42	50.0	1059 4	US-09-270-767-45764 Sequence 45764, A
8	41	48.8	400 4	US-09-252-991A-24981 Sequence 24981, A
9	41	48.8	545 3	US-09-269-731-8 Sequence 8, Appl
10	40	47.6	50 4	US-09-896-720-22 Sequence 22, Appl
11	40	47.6	102 4	US-09-248-796A-26528 Sequence 26528, A
12	40	47.6	112 3	US-09-134-001C-5057 Sequence 5057, Ap
13	40	47.6	159 2	US-08-829-110-1 Sequence 1, Appl
14	40	47.6	171 4	US-09-252-991A-30524 Sequence 30524, A
15	40	47.6	450 4	US-09-896-720-2 Sequence 2, Appl
16	40	47.6	452 4	US-09-270-767-43029 Sequence 43029, A
17	40	47.6	492 3	US-09-724-864-39 Sequence 39, Appl
18	39	46.4	67 4	US-09-270-767-38270 Sequence 38270, A
19	39	46.4	67 4	US-09-270-767-53487 Sequence 53487, A
20	39	46.4	178 4	US-09-328-352-4296 Sequence 4296, Ap
21	39	46.4	324 4	US-09-248-796A-20599 Sequence 20599, A
22	39	46.4	342 4	US-09-252-991A-32728 Patent No. 5352575
23	39	46.4	350 6	US-09-270-767-42648 Sequence 42648, A
24	39	46.4	423 4	US-09-270-767-42648 Sequence 42648, App
25	38	45.2	64 4	US-09-732-210-968 Sequence 968, A
26	38	45.2	71 4	US-09-248-796A-24429 Sequence 24429, A
27	38	45.2	118 3	US-09-199-637A-279 Sequence 279, App

RESULT 1
US-09-248-796A-19247
; Sequence 19247, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19247
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19247

ALIGNMENTS

28	38	45.2	157	4	US-09-248-796A-21632	Sequence 21632, A
29	38	45.2	170	4	US-09-252-991A-16827	Sequence 16827, A
30	38	45.2	178	4	US-09-270-767-43724	Sequence 43724, A
31	38	45.2	181	4	US-09-492-308A-2	Sequence 2, Appli
32	38	45.2	196	4	US-09-328-352-4778	Sequence 4778, Ap
33	38	45.2	233	4	US-09-107-532A-6776	Sequence 6776, Ap
34	38	45.2	276	4	US-09-543-681A-7430	Sequence 7430, Ap
35	38	45.2	288	3	US-09-492-985-2	Sequence 2, Appli
36	38	45.2	294	4	US-09-270-767-41531	Sequence 41531, A
37	38	45.2	323	3	US-08-532-896-2	Sequence 2, Appli
38	38	45.2	323	4	US-09-702-705-1821	Sequence 1821, Ap
39	38	45.2	323	4	US-09-736-457-1821	Sequence 1821, Ap
40	38	45.2	323	4	US-09-671-325-1821	Sequence 1821, Ap
41	38	45.2	364	4	US-09-643-597-172	Sequence 172, App
42	38	45.2	364	4	US-09-480-884A-172	Sequence 172, App
43	38	45.2	364	4	US-09-702-705-783	Sequence 783, App
44	38	45.2	364	4	US-09-736-457-783	Sequence 783, App
45	38	45.2	364	4	US-09-542-615A-172	Sequence 172, App

Query Match 51.2%; Score 43; DB 4; Length 461;
Best Local Similarity 46.2%; Pred. No. 38;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Qy 4 KSKRNPNALTPP 16
Db 12 KRRKNPSTTTPP 24
|.:.:|:|:|

RESULT 2
US-09-790-045-14
; Sequence 14, Application US/097900045
; Patent No. 6492510
; GENERAL INFORMATION:
; APPLICANT: Hasebe, Akira
; APPLICANT: Tsuchiya, Kenichi
; APPLICANT: Horita, Mitsuo
; TITLE OF INVENTION: Insertion Sequence Element Derived From Ralstonia Solanacearum
; Patent No. 6492510
; FILE REFERENCE: NANP108US
; CURRENT APPLICATION NUMBER: US/09/790,045
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Ralstonia solanacearum


```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45764
; LENGTH: 1059
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-45764

Query Match      50.0%; Score 42; DB 4; Length 1059;
Best Local Similarity 70.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      7 KRNPANLTPP 16
Db      456 RRRPQLNLTTP 465

RESULT 8
US-09-252-991A-24981
; Sequence 24981, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24981
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24981

Query Match      48.8%; Score 41; DB 4; Length 400;
Best Local Similarity 46.7%; Pred. No. 69;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      2 LAKSKENPANLTPP 16
Db      127 LARRQRDPAGAQP 141

RESULT 9
US-09-269-731-8
; Sequence 8, Application US/09269731
; Patent No. 633185
; GENERAL INFORMATION:
; APPLICANT: BARBEYRON, Tristan
; APPLICANT: FOTIN, Philippe
; APPLICANT: RICHARD, Christophe
; APPLICANT: HENRISSAT, Bernard
; APPLICANT: YVIN, Jean-Claude
; APPLICANT: KLOAREG, Bernard
; TITLE OF INVENTION: Glycolyse hydrolase genes and their
; TITLE OF INVENTION: use for producing enzymes for the biodegradation of
; TITLE OF INVENTION: carrageenans
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DENNISON, MESEROLE, SCHEINER & SCHULTZ
; STREET: 612 Crystal Square 4, 1745 Jefferson Davis
; STREET: Highway
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/269,731
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR97/01768
; FILING DATE: 06-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 96 12204
; FILING DATE: 07-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: IRA SCHULTZ
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 412-1155
; TELEFAX: (703) 412-1161
; INFORMATION FOR SEQ ID NO: 8:
; LENGTH: 545 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-269-731-8

Query Match      48.8%; Score 41; DB 3; Length 545;
Best Local Similarity 50.0%; Pred. No. 95;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      1 GLAKSKENPANLTPP 16
Db      87 GIAELTMRHNNANLTP 102

RESULT 10
US-09-896-720-22
; Sequence 22, Application US/09896720
; Patent No. 6750057
; GENERAL INFORMATION:
; APPLICANT: SONG, WEN-YUAN
; APPLICANT: PI, LI-YA
; TITLE OF INVENTION: UBIQUITIN LIGASE
; FILE REFERENCE: 5853-173
; CURRENT APPLICATION NUMBER: US/09/896,720
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/215,049
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-896-720-22

Query Match      47.6%; Score 40; DB 4; Length 50;
Best Local Similarity 63.6%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      6 SKRNPANLTPP 16
Db      32 NKPNTTLTPP 42

RESULT 11
US-09-248-796A-26528
; Sequence 26528, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
```

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; CURRENT FILING DATE: 1998-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; PRIOR FILING DATE: 1998-08-13
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 26528
 ; LENGTH: 102
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 ; US-09-248-796A-26528

Query Match 47.6%; Score 40; DB 4; Length 102;
 Best Local Similarity 46.2%; Pred. No. 25;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 3 AKSKRNPNLTPP 15
 Db 3 SKERQTNPSNTP 15

RESULT 12
 US-09-134-001C-5057
 ; Sequence 5057, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 5057
 ; LENGTH: 112
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-5057

Query Match 47.6%; Score 40; DB 3; Length 112;
 Best Local Similarity 50.0%; Pred. No. 27;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 AKSKRNPNLTPP 16
 Db 25 AKKNMPPFGFNPP 38

RESULT 13
 US-08-829-110-1
 ; Sequence 1, Application US/08829110
 ; Patent No. 5882890
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Goli, Surya K.
 ; TITLE OF INVENTION: NOVEL REGULATORS OF G-PROTEIN
 ; TITLE OF INVENTION: SIGNALING
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Icyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/829,110
 ; FILING DATE: Filed Herewith
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0259 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 159 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: ADENINE01
 ; CLONE: 158909
 ; US-08-829-110-1

Query Match 47.6%; Score 40; DB 2; Length 159;
 Best Local Similarity 80.0%; Pred. No. 39;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 KSKRNPNLTP 14
 Db 15 KSKRPPSNLT 24

RESULT 14
 US-09-252-991A-30524
 ; Sequence 30524, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 30524
 ; LENGTH: 171
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-30524

Query Match 47.6%; Score 40; DB 4; Length 171;
 Best Local Similarity 50.0%; Pred. No. 42;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 AKSKRNPNLTPP 16
 Db 81 ACRSRNFGSPSP 94

RESULT 15
 US-09-896-720-2
 ; Sequence 2, Application US/09896720
 ; Patent No. 6750057

```
; GENERAL INFORMATION:
; APPLICANT: SONG, WEN-YUAN
; APPLICANT: PI, LI-YA
; TITLE OF INVENTION: UBIQUITIN LIGASE
; FILE REFERENCE: 5853-173
; CURRENT APPLICATION NUMBER: US/09/896,720
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/215,049
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (258)
; OTHER INFORMATION: Any amino acid
US-09-896-720-2

Query Match      47.6%; Score 40; DB 4; Length 450;
Best Local Similarity 63.6%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      6 SKRNPANLTPP 16
      :|||
Db      354 NKPNTTLTPP 364

Search completed: November 10, 2004, 13:44:10
Job time : 6.52677 secs
```

This Page Blank (uspto)

Sequence 169759,
Sequence 279854,
Sequence 127613,
Sequence 227000,
Sequence 181810,
Sequence 278969,
Sequence 14, Appl
Sequence 14, Appl
Sequence 14, Appl
Sequence 161558,
Sequence 65054, A
Sequence 69888, A
Sequence 8199, Ap
Sequence 12821,
Sequence 263705,
Sequence 207353,
Sequence 294759,
Sequence 220523,
Sequence 136603,
Sequence 135862,
Sequence 646, App
Sequence 212002,
Sequence 273102,
Sequence 8, Appli
Sequence 8, Appli
Sequence 8, Appli
Sequence 197986,
Sequence 22, Appl
Sequence 30932, A
Sequence 366586,
Sequence 190320,
Sequence 195287,

66 16 US-10-437-963-169759
134 17 US-10-425-115-279854
144 16 US-10-437-963-127613
156 17 US-10-425-115-227000
96 15 US-10-424-599-181810
127 17 US-10-425-115-278969
274 9 US-09-790-045-14
274 14 US-10-222-577-14
274 14 US-10-222-578-14
385 16 US-10-437-963-161558
572 15 US-10-282-122A-65054
572 15 US-10-282-122A-65054
1068 14 US-10-128-714-8199
51 17 US-10-425-115-192821
81 17 US-10-425-115-263705
41 48.8 81 17 US-10-424-599-207353
41 48.8 85 15 US-10-425-115-294759
41 48.8 95 17 US-10-425-115-220523
126 17 US-10-425-115-220523
183 16 US-10-437-963-135862
206 16 US-10-437-963-135862
268 16 US-10-408-765A-646
407 17 US-10-425-115-212002
456 15 US-10-424-599-273102
545 9 US-09-988-201-8
545 9 US-09-988-202-8
545 9 US-09-988-200-8
741 16 US-10-437-963-197986
50 9 US-09-896-720-22
53 14 US-10-029-386-30932
74 17 US-10-425-115-366586
82 17 US-10-425-115-190320
92 16 US-10-437-963-195287

ALIGNMENTS

RESULT 1
US-10-092-750-39
; Sequence 39, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-39

Query Match 100.0%; Score 84; DB 14; Length 16;
Best Local Similarity 100.0%; Pred. No. 4e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLAKSKRNPNALTPP 16
DB 1 GLAKSKRNPNALTPP 16
RESULT 2
US-10-264-049-3417
; Sequence 3417, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 10, 2004, 16:36:12 ; Search time 17.1883 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-39
Perfect score: 84
Sequence: 1 GLAKSKRNPNALTPP 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues
Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	100.0	16	14	US-10-092-750-39
2	62	73.8	23	15	US-10-864-049-3417
3	48	57.1	94	17	US-10-425-115-211273
4	47	56.0	143	17	US-10-425-115-228508
5	47	56.0	324	15	US-10-425-114-64388
6	47	56.0	446	17	US-10-425-115-250313
7	46	54.8	184	15	US-10-424-599-211116
8	46	54.8	538	17	US-10-425-115-191510
9	44.5	53.0	170	15	US-10-424-599-168814
10	44	52.4	109	17	US-10-425-115-253454
11	44	52.4	160	17	US-10-425-115-335386
12	44	52.4	247	17	US-10-425-115-306150
13	43.5	51.8	81	17	US-10-425-115-353683

APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 3417
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-3417

Query Match 73.8%; Score 62; DB 15; Length 23;
Best Local Similarity 75.0%; Pred. No. 0.015;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GLAKSKRNPNLTPP 16
Db 1 GKVKKKNSANLTPP 16

RESULT 3
US-10-425-115-211273
; Sequence 211273, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 211273
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_124277C.1.pep
US-10-425-115-211273

Query Match 57.1%; Score 48; DB 17; Length 94;
Best Local Similarity 60.0%; Pred. No. 9.1;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 LAKSKRNPNLTPP 16
Db 75 LFKEKKNGPLTPP 89

RESULT 4
US-10-425-115-228508
; Sequence 228508, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants

FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 228508
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_139992C.1.pep
US-10-425-115-228508

Query Match 56.0%; Score 47; DB 17; Length 143;
Best Local Similarity 81.8%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 SKRNPNLTPP 16
Db 97 SKRNPNLTPP 107

RESULT 5
US-10-425-114-64638
; Sequence 64638, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 64638
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73159D10_FLI.pep
US-10-425-114-64638

Query Match 56.0%; Score 47; DB 15; Length 324;
Best Local Similarity 81.8%; Pred. No. 46;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 SKRNPNLTPP 16
Db 234 SKRNPNLTPP 244

RESULT 6
US-10-425-115-250313
; Sequence 250313, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 250313
; LENGTH: 446
; TYPE: PRT

;; ORGANISM: Zea mays
;; FEATURE:
US-10-425-115-250313
Query Match 56.0%; Score 47; DB 17; Length 446;
Best Local Similarity 81.8%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 SKRNPANLTPP 16
DB 356 SKRNPANLTPP 366

RESULT 7
US-10-424-599-211116
; Sequence 211116, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 211116
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: unsure
; LOCATION: (1)..(184)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_32664C.1.pep
US-10-424-599-211116

Query Match 54.8%; Score 46; DB 15; Length 184;
Best Local Similarity 53.8%; Pred. No. 37;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 KSKRNPANLTPP 16
DB 24 KSKRNPANLTPP 36

RESULT 8
US-10-425-115-191510
; Sequence 191510, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 191510
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(538)
; OTHER INFORMATION: Clone ID: MRT4577_106236C.1.pep
US-10-425-115-191510

Query Match 54.8%; Score 46; DB 17; Length 538;
Best Local Similarity 53.3%; Pred. No. 11e-02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 LAKSKRNPANLTPP 16
DB 358 LAKSKRNPANLTPP 372

RESULT 9
US-10-424-599-168814
; Sequence 168814, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168814
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: unsure
; LOCATION: (1)..(170)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123454C.1.pep
US-10-424-599-168814

Query Match 53.0%; Score 44.5; DB 15; Length 170;
Best Local Similarity 62.5%; Pred. No. 58;
Matches 10; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 GLAKSKRNPANLTPP 16
DB 72 GLAKSKRNPANLTPP 86

RESULT 10
US-10-425-115-253454
; Sequence 253454, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 253454
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(109)
; OTHER INFORMATION: Clone ID: MRT4577_162731C.1.pep
US-10-425-115-253454

Query Match 52.4%; Score 44; DB 17; Length 109;
Best Local Similarity 52.9%; Pred. No. 44;
Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 GLAKKS--XRNPNALTPP 15

Db 66 GIKKKTGFKNPVNCTP 82

```

RESULT 11
US-10-425-115-335386
; Sequence 335386, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 335386
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(160)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_68982C.1.pep
US-10-425-115-335386

```

```

US-10-425-115-353683
; Sequence 353683, Application US/10425115
; Publication No. US2004021472A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 353683
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MET4577_85733C.1.psp
US-10-425-115-353683

Query Match          51.8%;      Score 43.5;   DB 17;    Length 81;
Best Local Similarity 56.2%;      Pred. No. 39;
Matches           9; Conservative         4; Mismatches       2; Indels        1; Gaps        1;

Qy     1 GLAKSKRNPNALTTPP 16
Db     7 GPIPKSOR-PLSLTTP 21
      |:|::|||::|||
RESULT'14
US-10-437-963-169759
; Sequence 169759, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 169759
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_68148C.1.pep
US-10-437-963-169759

```

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 279854
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_186806C.1.pep
US-10-425-115-279854

Query Match      51.2%; Score 43; DB 17; Length 134;
Best Local Similarity 50.0%; Pred.No. 77;
Matches      8; Conservative      2; Mismatches      6; Indels      0; Gaps      0;

QY      1 GLAKKSKENPANTLTPP 16
      |:|||||:
Db      118 GVAKKKKKKKKTTPP 133

```

Search completed: November 11, 2004, 01:28:21
Job time : 18.2383 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 3.56477 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-39

Perfect score: 84

Sequence: 1 GLAKSKRNPNLTPP 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.79:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	52.4	306	2 S67131	hypothetical prote
2	43	51.2	1171	2 T31635	hypothetical prote
3	42	50.0	409	2 S70704	carbon catabolite
4	42	50.0	419	2 T10652	hypothetical prote
5	42	50.0	572	2 F81866	arginine-tRNA liga
6	42	50.0	572	2 F81075	arginyl-tRNA synth
7	42	50.0	1032	2 T14124	neural zinc finger
8	41.5	49.4	726	2 S67044	NF11 protein - yea
9	41	48.8	268	2 G02133	holocytochrome-c s
10	41	48.8	269	2 T26823	hypothetical prote
11	41	48.8	280	2 D86193	hypothetical prote
12	41	48.8	281	2 T45926	hypothetical prote
13	41	48.8	565	2 H69363	hypothetical prote
14	41	48.8	629	2 B82545	two-component syst
15	41	48.8	777	2 S30271	pyocin Ap41 large
16	41	48.8	843	2 S33442	EF protein - Strept
17	41	48.8	1822	2 S33441	BF protein - Strept
18	40	47.6	116	2 E72622	hypothetical prote
19	40	47.6	164	2 A85093	hypothetical prote
20	40	47.6	212	2 A84971	dTMP kinase (EC 2.
21	40	47.6	346	2 AC3108	hypothetical prote
22	40	47.6	371	2 A98179	ABC transporter, A
23	40	47.6	374	1 LYX1Y	beta-lytic metallo
24	40	47.6	567	2 AG2008	hypothetical prote
25	40	47.6	683	2 T34103	hypothetical prote
26	40	47.6	697	2 C75525	molybdopterin oxid
27	40	47.6	720	2 T47648	ABC transporter-11
28	40	47.6	1122	2 T14180	exit protein - Myc
29	40	47.6	2252	2 S06188	genome polyprotein

30 39.5 47.0 780 2 T50315
31 39 46.4 118 2 E75077
32 39 46.4 184 2 T26262
33 39 46.4 277 2 AD3322
34 39 46.4 298 2 T47760
35 39 46.4 350 1 VGBE63
36 39 46.4 415 2 E71677
37 39 46.4 497 2 D84778
38 39 46.4 586 2 T36252
39 39 46.4 644 2 S39356
40 39 46.4 676 2 S61977
41 39 46.4 688 2 AB1257
42 39 46.4 696 2 T27402
43 39 46.4 707 2 S57157
44 39 46.4 786 2 E86268
45 39 46.4 929 2 T52517

ALIGNMENTS

RESULT 1

S67131 hypothetical protein YOR238w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein OS237

C:Species: Saccharomyces cerevisiae

C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004

C:Accession: S67131

R:Boyer, J.; Fairhead, C.; Gaillon, L.; Galisson, F.; Michaux, G.; Thierry, A.; Dujon, B.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67104

A:Accession: S67131

A:Molecule type: DNA

A:Residues: 1-306 <BO>

A:Cross-references: UNIPROT:Q08634; EMBL:Z75146; NID:g1420545; GSPDB:GN0001

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YOR238w

A:Cross-references: SGD:S0005764

A:Map position: 15R

Query Match 52.4%; Score 44; DB 2; Length 306;

Best Local Similarity 60.0%; Pred. No. 13;

Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 LAKSKRNPNLTPP 16

Db 253 LTKSKSRNPNRTAP 267

RESULT 2

T31635 hypothetical protein Y57A10A.m - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000

C:Accession: T31635

R:Smyle, R.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z21048

A:Accession: T31635

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1171 <WIL>

A:Cross-references: EMBL:AL117195; PIDN:CAB55018.1; CESP:Y57A10A.m

A:Experimental source: clone Y57A10A

C:Genetics:

A:Gene: CESP:Y57A10A.m

A:Introns: 106/3; 257/3; 286/1; 405/2; 504/2; 687/2; 811/1; 945/2; 1045/3; 1121/3

C:Superfamily: DNA-directed DNA polymerase, mitochondrial

Query Match 51.2%; Score 43; DB 2; Length 1171;

Best Local Similarity 66.7%; Pred. No. 79;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

QY      3 AKKSKRNPNANIT 14
DB      455 SKKSKRNPNKST 466

RESULT 3
S70704
Carbon catabolite repressor creI - fungus (Trichoderma harzianum)
N:Alternate names: glucose repressor creI
C:Species: Trichoderma harzianum
C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 09-Jul-2004
C:Accession: S70704; S77920; S73106
R:Illmen, M.; Thrane, C.; Penttilae, M.
Mol. Gen. Genet. 251, 451-460, 1996
A:Title: The glucose repressor gene creI of Trichoderma: isolation and expression of a
A:Reference number: S70703; MUID:96285569; PMID:8709949
A:Accession: S70704
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-409 <ILM1>
A:Cross-references: UNIPROT:Q99005; EMBL:X95369
A:Experimental source: strain T3
A:Accession: S77920
A:Molecule type: mRNA
A:Residues: 1-409 <ILM2>
A:Cross-references: EMBL:X95369
A:Experimental source: strain T3
R:Illmen, M.
submitted to the EMBL Data Library, January 1996
A:Reference number: S73106
A:Accession: S73106
A:Molecule type: DNA
A:Residues: 1-268, 'S', 270-409 <ILM>
A:Cross-references: EMBL:X95369; NID:g1177688; PID:g1177689
A:Experimental source: strain T3
C:Genetics:
A:Gene: creI
C:Keywords: DNA binding; repressor; transcription regulation; zinc finger

Query Match      50.0%; Score 42; DB 2; Length 409;
Best Local Similarity 57.1%; Pred. No. 38;
Matches      8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      3 AKKSKRNPNANITPP 16
DB      262 AKRSRPNPNSTAP 275

RESULT 4
T10652
Hypothetical protein T5F17.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T10652
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T10652
A:Molecule type: DNA
A:Residues: 1-419 <BEV>
A:Cross-references: UNIPROT:Q9MGH3; EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.30
A:Experimental source: cultivar Columbia; BAC clone T5F17
C:Genetics:
A:Gene: ATSP:T5F17.30
A:Map position: 4
A:Introns: 97/1; 243/3; 328/3
C:Superfamily: Arabidopsis thaliana hypothetical protein F17D16.20

Query Match      50.0%; Score 42; DB 2; Length 419;
Best Local Similarity 50.0%; Pred. No. 39;
Matches      8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      1 GLAKSKRNPNANITPP 16
DB      16 GTAVKRNRMPSLSPP 31

RESULT 5
F81866
arginine-tRNA ligase (EC 6.1.1.19) NMA1707 [imported] - Neisseria meningitidis (strain Z:
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C:Accession: F81866
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:2022556; PMID:10761919
A:Accession: F81866
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-572 <PAR>
A:Cross-references: UNIPROT:Q9JTM7; GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB8493;
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: args; NMA1707
C:Superfamily: arginine-tRNA ligase
C:Keywords: ligase

Query Match      50.0%; Score 42; DB 2; Length 572;
Best Local Similarity 61.5%; Pred. No. 54;
Matches      8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      1 GLAKSKRNPNANIT 13
DB      45 GAAXKAKQNPREL 57

RESULT 6
F81075
arginyl-tRNA synthetase NMB1506 [imported] - Neisseria meningitidis (strain MC58 serogro
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: F81075
R:Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: F81075
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-572 <TER>
A:Cross-references: UNIPROT:Q9JYM8; GB:AE002500; GB:AE002098; NID:g7226744; PIDN:AAF4186;
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1506
C:Superfamily: arginine-tRNA ligase

Query Match      50.0%; Score 42; DB 2; Length 572;
Best Local Similarity 61.5%; Pred. No. 54;
Matches      8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      1 GLAKSKRNPNANIT 13
DB      45 GAAXKAKQNPREL 57

RESULT 7
T14124
neural zinc finger factor 3 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

```

C;Accession: T14124
 R;Yee, K.S.Y.; Yu, V.C.
 J. Biol. Chem. 273, 5366-5374, 1998
 A;Title: Isolation and characterization of a novel member of the neural zinc finger factor
 A;Reference number: Z17882; MUID:98148091; PMID:9478997
 A;Accession: T14124
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1032 <YEE>
 A;Cross-references: UNIPROT:Q9QX27; EMBL:AF031942; NID:g2914750; PID:g2914751; PIDN:AAAC
 C;Genetics:
 A;Note: NZF-3
 C;Function:
 A;Description: confers repression on the basal activity of promoters containing the cons
 C;Keywords: DNA binding; transcription factor; zinc finger

Query Match 50.0%; Score 42; DB 2; Length 1032;
 Best Local Similarity 61.5%; Pred. No. 1e+02; 4; Indels 0; Gaps 0;
 Matches 8; Conservative 1; Mismatches 0; Gaps 0;

Qy 3 AKKSKRNPNALTP 15
 Db 829 AKRQENFLNGTP 841
 ||: ||||| |||
 ||: ||||| |||

RESULT 8
 S67044
 NF11 protein - yeast (Saccharomyces cerevisiae)
 N;Alternate names: protein Q3551; protein YOR156c
 C;Species: Saccharomyces cerevisiae
 C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
 A;Accession: S67044; S59360
 R;Bordome, R.; Canasess, A.; Madania, A.; Martin, R.P.; Poch, O.; Tarassov, I.A.; Winsc
 submitted to the Protein Sequence Database, July 1996
 A;Reference number: S67032
 A;Accession: S67044
 A;Molecule type: DNA
 A;Residues: 1-726 <BOR>
 A;Cross-references: UNIPROT:Q12216; EMBL:Z75064; NID:g1420388; PID:e252042; PID:g1420389
 A;Experimental source: strain S288C
 R;Pringle, J.R.
 submitted to the EMBL Data Library, August 1995
 A;Description: S. cerevisiae two-hybrid interactor with the C-terminus of Cdc12p.
 A;Reference number: S59360
 A;Accession: S59360
 A;Molecule type: DNA
 A;Residues: 1-102, 'DKLIIT', 109, 'ILY', 113-560, 'T', 562-726 <PRI>
 C;Genetics:
 A;Gene: SGD:NFI1
 A;Cross-references: SGD:S0005682; MIPS:YOR156c
 A;Map position: 15R

Query Match 49.4%; Score 41.5; DB 2; Length 726;
 Best Local Similarity 62.5%; Pred. No. 84;
 Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

Qy 1 GLAKKS-KRNPNALTP 15
 Db 244 GLKQNGTGNPNALTP 259
 ||: ||||| |||
 ||: ||||| |||

RESULT 9
 G02133
 holocytochrome-c synthase (EC 4.4.1.17) - human
 C;Species: Homo sapiens (man)
 C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
 A;Accession: G02133
 R;Schaefer, L.; Ballabio, A.; Zoghbi, H.Y.
 submitted to the EMBL Data Library, September 1995
 A;Reference number: H00824
 A;Accession: G02133
 A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA
 A;Residues: 1-268 <SCH>
 A;Cross-references: UNIPROT:P53701; EMBL:U36787; NID:g1209634; PIDN:AAB15007.1; PID:g120
 C;Genetics:
 A;Gene: GDB:HCCS; CCHL
 A;Cross-references: GDB:636832; OMIM:300056
 A;Map position: Xp22-Xp22
 C;Keywords: carbon-sulfur lyase; mitochondrion

Query Match 48.8%; Score 41; DB 2; Length 268;
 Best Local Similarity 50.0%; Pred. No. 36;
 Matches 9; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

Qy 1 GLAKSKRN--PANLTPP 16
 Db 70 GTAAENKENLDPNLMP 87
 ||: ||||| |||
 ||: ||||| |||

RESULT 10
 T26823
 hypothetical protein Y43CSA.5 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C;Accession: T26823
 R;White, S.
 submitted to the EMBL Data Library, June 1998
 A;Reference number: Z20272
 A;Accession: T26823
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-269 <WIL>
 A;Cross-references: UNIPROT:Q9XXF3; EMBL:AL023838; PIDN:CAAL19501.1; GSPDB:GNO0022; CESP:
 A;Experimental source: clone Y43CSA
 C;Genetics:
 A;Gene: CESP:Y43CSA.5
 A;Map position: 4
 A;Introns: 9/3; 47/2; 85/2; 227/3

Query Match 48.8%; Score 41; DB 2; Length 269;
 Best Local Similarity 46.7%; Pred. No. 36;
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 2 LAKKSKRNPNALTPP 16
 Db 237 LAKKQKSDGSVSPP 251
 ||||| ||: |||||
 ||||| ||: |||||

RESULT 11
 D86193
 hypothetical protein [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C;Accession: D86193
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A;Reference number: A86141; MUID:21016719; PMID:11130712
 A;Accession: D86193
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-280 <STO>
 A;Cross-references: UNIPROT:Q9MA40; GB:AE005172; NID:g6850315; PIDN:AAF29392.1; GSPDB:GN
 C;Genetics:
 A;Map position: 1
 C;Superfamily: Arabidopsis thaliana hypothetical protein F5K20.160

Query Match 48.8%; Score 41; DB 2; Length 280;
 Best Local Similarity 46.7%; Pred. No. 38;
 Matches 7; Conservative 4; Mismatches 0; Gaps 0;

QY 2 LAKSKRNPNANLTPP 16
 :|||:|||||
 Db 1 MASASKONPSSSKPP 15

RESULT 12
 T45926
 Hypothetical protein F5K20.160 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
 C:Accession: T45926
 R:Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.P.X.
 submitted to the Protein Sequence Database, January 2000
 Reference number: Z23017
 A:Accession: T45926
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-281 <MON>
 A:Cross-references: UNIPROT:Q9M340; EMBL:AL132960
 A:Experimental source: cultivar Columbia; BAC clone F5K20
 C:Genetics:
 A:Map position: 3
 A:Introns: 194/2
 A:Note: F5K20.160
 C:Superfamily: Arabidopsis thaliana hypothetical protein F5K20.160

Query Match 48.8%; Score 41; DB 2; Length 281;
 Best Local Similarity 50.0%; Pred. No. 38;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 AKKSKRNPNANLTPP 16
 |||:|||||
 Db 5 AKNTQNPFSLKPP 18

RESULT 13
 H69363
 Hypothetical protein AF0912 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
 C:Accession: H69363
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.P.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
 A:Reference number: A69250; PMID:98049343; PMID:9389475
 A:Accession: H69363
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-565 <KLE>
 A:Cross-references: UNIPROT:Q29350; GB:AE001041; GB:AE000782; NID:G2689364; PIDN:AA39034

Query Match 48.8%; Score 41; DB 2; Length 565;
 Best Local Similarity 57.1%; Pred. No. 79;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 AKKSKRNPNANLTPP 16
 |||:|||||
 Db 511 AKASQQTAPATPP 524

RESULT 14
 B82545
 two-component system, sensor protein XP2546 [imported] - Xylella fastidiosa (strain 9a5c
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

C:Accession: B82545
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 405, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; PMID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: B82545
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-629 <SIM>
 A:Cross-references: UNIPROT:Q9PAH2; GB:AE004061; GB:AE003849; NID:G9107747; PIDN:AAF85343
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; AJ
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2546

Query Match 48.8%; Score 41; DB 2; Length 629;
 Best Local Similarity 53.8%; Pred. No. 88;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 GLAKSKRNPNANL 13
 |||:|||||
 Db 466 GLARRERANPENL 478

RESULT 15
 S30271
 pyocin AP41 large chain - Pseudomonas aeruginosa transposon TnAP41
 C:Species: Pseudomonas aeruginosa
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
 C:Accession: S30271; A47060; S27605
 R:Sano, Y.; Kageyama, M.
 Mol. Gen. Genet. 237, 161-170, 1993
 A:Title: A novel transposon-like structure carries the genes for pyocin AP41, a Pseudom
 A:Reference number: S30271; PMID:93204890; PMID:8384291
 A:Accession: S30271
 A:Molecule type: DNA
 A:Residues: 1-777 <SAN>
 A:Cross-references: UNIPROT:Q51502; GB:D12705; NID:G216903; PIDN:BA02196.1; PID:G216904
 R:Sano, Y.
 J. Bacteriol. 175, 912-915, 1993
 A:Title: The inherent DNase of pyocin AP41 causes breakdown of chromosomal DNA.
 A:Reference number: A47060; PMID:93139066; PMID:8423163
 A:Accession: A47060
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 642-762 <SA2>
 C:Superfamily: pyocin AP41 large chain
 C:Keywords: bacteriocin

Query Match 48.8%; Score 41; DB 2; Length 777;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 KXSKRNPNANLTP 15
 |||:|||||
 Db 106 KXKRNPNANLTP 117

Search completed: November 10, 2004, 13:40:47

Job time : 4.56477 secs



This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 19.5371 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-39

Sequence: 1 GLAKSKRNPANLTPP 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	47	56.0	623	2	Q8AW61	Q8AW61	brachydanio
2	47	56.0	716	2	Q6C2U3	Q6C2U3	varrowia li
3	46	54.8	199	2	Q75CN1	Q75CN1	ashbya goss
4	46	54.8	199	2	AA51116	AA51116	ashbya goss
5	45	53.6	696	2	Q7VB38	Q7VB38	prochloroco
6	45	53.6	1943	2	Q6SLC1	Q6SLC1	cochliobol
7	45	53.6	1943	2	AA229905	AA229905	cochliobol
8	44	52.4	306	2	Q08634	Q08634	saccharomyc
9	44	52.4	453	2	Q86KC9	Q86KC9	dictyosteli
10	44	52.4	688	2	Q6BNW6	Q6BNW6	debaromyce
11	44	52.4	3539	2	Q9VZ30	Q9VZ30	drosophila
12	44	52.4	3539	2	AA48000	AA48000	drosophila
13	44	52.4	4793	2	Q7Z8F4	Q7Z8F4	emerice
14	43	51.2	167	2	Q74N70	Q74N70	nanosarchaeu
15	43	51.2	167	2	AA38927	AA38927	nanosarchaeu
16	43	51.2	172	2	Q6Z2D7	Q6Z2D7	oryza sativ
17	43	51.2	172	2	BAD03518	BAD03518	oryza sativ
18	43	51.2	389	2	Q759V2	Q759V2	ashbya goss
19	43	51.2	389	2	AA52091	AA52091	ashbya goss
20	43	51.2	428	2	Q7RZ28	Q7RZ28	neutrospora
21	43	51.2	662	2	Q9L4V1	Q9L4V1	streptomyce
22	43	51.2	2805	2	Q7Q3N3	Q7Q3N3	anopheles g
23	42	50.0	170	2	Q7U224	Q7U224	rhodospir
24	42	50.0	237	2	Q31466	Q31466	morone saxa
25	42	50.0	237	2	Q31471	Q31471	morone saxa
26	42	50.0	258	2	Q72RC0	Q72RC0	leptospi
27	42	50.0	258	2	Q8F4F0	Q8F4F0	leptospi
28	42	50.0	258	2	AA570414	AA570414	leptospi
29	42	50.0	274	2	Q05546	Q05546	gluconobact
30	42	50.0	274	2	Q9FA37	Q9FA37	raistonia s
31	42	50.0	274	2	Q8XU19	Q8XU19	raistonia s

32	42	50.0	277	2	Q88EG7	Q88EG7	pseudomonas
33	42	50.0	351	2	Q7Q2T0	Q7Q2T0	anopheles g
34	42	50.0	414	2	Q7BJJ0	Q7BJJ0	anopheles g
35	42	50.0	419	2	Q9M0H3	Q9M0H3	arabidopsis
36	42	50.0	425	2	Q73MJ4	Q73MJ4	treponema d
37	42	50.0	425	2	AA512031	AA512031	treponema
38	42	50.0	572	1	SYR NEIMA	SYR NEIMA	neisseria m
39	42	50.0	572	1	SYR NEIMB	SYR NEIMB	neisseria m
40	42	50.0	872	2	Q8IHR3	Q8IHR3	plasmidium
41	42	50.0	1014	2	Q96PY4	Q96PY4	homo sapien
42	42	50.0	1032	2	Q90X27	Q90X27	rattus norv
43	42	50.0	1059	2	Q9VW51	Q9VW51	drosophila
44	42	50.0	1059	2	AA49098	AA49098	drosophila
45	42	50.0	1060	2	Q95VE6	Q95VE6	drosophila

ALIGNMENTS

RESULT 1	Q8AW61	PRELIMINARY;	PRT;	623 AA.
ID	Q8AW61			
AC	Q8AW61			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	SI:BAC6CSE.1 (Novel protein similar to human procadherin 15 (PCDH15))			
DE	(Fragment).			
OS	Name=BAC7C8S.1;			
OS	Brachydanio rerio (Zebrafish) (Danio rerio).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;			
OC	Cyprinidae; Danio.			
OX	NCBI_TaxID=7955;			
RN	[1]			
RP	Lloyd D.;			
RA	Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.			
RL	-1- SIMILARITY: Contains 4 cadherin domains.			
CC	EMBL; AL592062; CAD5120.1; -.			
DR	HSSP; P15116; INCU.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005509; F:calcium ion binding; IEA.			
DR	GO; GO:0007156; P:phospholipid cell adhesion; IEA.			
DR	InterPro; IPR002126; Cadherin.			
DR	Pfam; PF00028; Cadherin; 4.			
DR	PRINTS; PR00205; CADHERIN.			
DR	SMART; SM00112; CA; 5.			
DR	PROSITE; PS00232; CADHERIN_1; 2.			
DR	PROSITE; PS00268; CADHERIN_2; 4.			
KW	Calcium; Calcium-binding.			
FT	NON_TER 1			
FT	NON_TER 623 623			
SQ	SEQUENCE 623 AA; 69282 MW; 70BC5E2D0F8CFC3A CRC64;			

Query Match 56.0%; Score 47; DB 2; Length 623;
Best Local Similarity 60.0%; Pred. No. 39;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY	2 LAKSKRNPANLTPP 16
DB	236 LTDPKLNPLNTPP 250

RESULT 2

ID	Q6C2U3	PRELIMINARY;	PRT;	716 AA.
AC	Q6C2U3			
DT	01-OCT-2004 (TrEMBLrel. 28, Created)			
DT	01-OCT-2004 (TrEMBLrel. 28, Last sequence update)			
DT	01-OCT-2004 (TrEMBLrel. 28, Last annotation update)			
DE	Similar to KLA0D16115g Kluyveromyces lactis IPF 4333.1.			
GN	ORFNames=YAL10F05126g;			


```

DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01757; Acyl_transf_3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Acyltransferase; Complete proteome; Transferase.
SQ SEQUENCE 696 AA; 79572 MW; 628FF962C9072747 CRC64;

Query Match 53.6%; Score 45; DB 2; Length 696;
Best Local Similarity 69.2%; Pred. No. 97;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GLAKSKRNPNALTPP 13
Db 164 GFAXKSKNGPNL 176

RESULT 6
Q6SLC1 PRELIMINARY; PRT; 1943 AA.
AC Q6SLC1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Putative response regulator receiver RIM15p.
DE Name=RIM15;
GN Cochlidiobolus heterostrophus (Drechslera maydis).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochlidiobolus.
OX NCBI_TaxID=5016;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=C4;
RX PubMed=14665450;
RA Catlett N.L., Yoder O.C., Turgeon B.G.;
RT "Whole-genome analysis of two-component signal transduction genes in
RT fungal pathogens.";
RL Eukaryotic cell 2:1151-1161(2003).
DR EMBL; AY456029; AAR29905.1; -.
SQ SEQUENCE 1943 AA; 212673 MW; 8F8485E22FFC3A45 CRC64;

Query Match 53.6%; Score 45; DB 2; Length 1943;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GLAKSKRNPNALTPP 16
Db 218 GLAEQSATDPGNCAPP 233

RESULT 7
Q6SLC1 PRELIMINARY; PRT; 1943 AA.
AC Q6SLC1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Putative response regulator receiver RIM15p.
DE Name=RIM15;
GN Cochlidiobolus heterostrophus (Drechslera maydis).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochlidiobolus.
OX NCBI_TaxID=5016;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=C4;
RX PubMed=14665450;
RA Catlett N.L., Yoder O.C., Turgeon B.G.;
RT "Whole-genome analysis of two-component signal transduction genes in
RT fungal pathogens.";
RL Eukaryotic Cell 2:1151-1161(2003).
DR EMBL; AY456029; AAR29905.1; -.
SQ SEQUENCE 1943 AA; 212673 MW; 8F8485E22FFC3A45 CRC64;

Query Match 53.6%; Score 45; DB 2; Length 1943;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GLAKSKRNPNALTPP 16
Db 218 GLAEQSATDPGNCAPP 233

RESULT 8
Q08634 PRELIMINARY; PRT; 306 AA.
AC Q08634;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE S.cerevisiae chromosome XV reading frame ORF YOR238w.
DE ORFNames=YOR238W;
GN Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1];
RP SEQUENCE FROM N.A.
RA Boyer J., Fairhead C., Gaillon L., Galisson F., Michaux G.,
RA Thierry A., Dujon B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2];
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z75146; CAA99459.1; -.
DR PIR; S67131; S67131.
DR SGD; S0005764; YOR238W.
DR GO; GO:0005737; Cytoplasm; IDA.
SQ SEQUENCE 306 AA; 35591 MW; 6CA44BEDC9D7E4C CRC64;

Query Match 52.4%; Score 44; DB 2; Length 306;
Best Local Similarity 60.0%; Pred. No. 59;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 LAKSKRNPNALTPP 16
Db 253 LTKGSRNPFNRTAP 267

RESULT 9
Q86KC9 PRELIMINARY; PRT; 453 AA.
ID Q86KC9
AC Q86KC9;

```

DT 01-JUN-2003 (TReMBLrel. 24, Created)
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE=22092622; PubMed=12097910;
 RA Gloeckner G., Eichinger L., Safranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tungal B., Cox B., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
 RL Nature 418:79-85(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Baumgart C.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC116956; AAC01128.1; -;
 DR InterPro; IPR008615; FNIP.
 DR Pfam; PF05725; FNIP; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 453 AA; 51421 MW; 5C79A3D45C13E1CD CRC64;
 Query Match 52.4%; Score 44; DB 2; Length 453;
 Best Local Similarity 53.3%; Pred. No. 90;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 2 LAKSKRNPNLTPP 16
 DB 105 LNEKSLKPLNIYPP 119
 RESULT 10
 Q6BNW6 PRELIMINARY; PRT; 688 AA.
 AC Q6BNW6;
 DT 01-OCT-2004 (TReMBLrel. 28, Created)
 DT 01-OCT-2004 (TReMBLrel. 28, Last sequence update)
 DT 01-OCT-2004 (TReMBLrel. 28, Last annotation update)
 DE Chromosome E of strain CBS767 of Debaryomyces hansenii.
 GN ORFNames=DEHA0E19404g;
 OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
 OX NCBI_TaxID=4959;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS767;
 RG GENOLEVURES;
 RA Duion B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
 RA Goffard N., Frangeul L., Aigle M., Anhouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
 RA Boirame A., Boyer J., Cattolico L., Confalonieri F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Janiaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul B., Lemaire M., Lesur I., Va L., Muller H.,
 RA Niclaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenz S., Petier S., Richard G.F., Straub M.L., Suleau A.,
 RA Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RA Wincker P., Souciet J.L.;
 RT "Genome evolution in yeasts.";
 RL Nature 430:35-44(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS767;
 RA Genoscope;

RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; CR382137; CAG88371.1; -;
 SQ SEQUENCE 688 AA; 77169 MW; 23B2BBB3F07BBD71 CRC64;
 Query Match 52.4%; Score 44; DB 2; Length 688;
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LAKSKRNPNAL 13
 DB 272 LSKRIKPNAL 283
 RESULT 11
 Q9VZ30 PRELIMINARY; PRT; 3539 AA.
 ID Q9VZ30;
 AC Q9VZ30;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE CG1122-PA.
 GN ORFNames=CG11122;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=20196006; PubMed=10731132;
 RX Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 April J.F., Agbayani A., An H.J., Andrews-Frannkoch C., Baldwin D.,
 Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.W.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Fessler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gortell J.H., Gu Z., Guan P., Harris K.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusakern D.R., Pacle J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reiner K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.Y., Wasarman D.A., Weinstock G.M., Weissenbach J.,
 Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celisner S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,


```

RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
RN [5]
RN SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RN SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003485; AAF48000.3; -.
DR FLYBase; PBgn0030266; CG11122.
SQ SEQUENCE 3539 AA; 384994 MW; 0BD4A2997D18E0C1 CRC64;

Query Match 52.4%; Score 44; DB 2; Length 3539;
Best Local Similarity 53.3%; Pred. No. 8.4e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 LAKSKRNPANLTPP 16
DB 1502 LSKSKSRKPSVPFP 1516
|:|||||:|:|:|

RESULT 13
Q7Z8P4 PRELIMINARY; PRT; 4793 AA.
AC Q7Z8P4;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DE 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Peptide synthetase.
GN Name=sidC;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Peizomycotina; Eurotiomycetes;
OC Eutotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22713765; PubMed=12828635;
RA Eiselndle M., Oberegger H., Zadra I., Haas H.;
RT "The siderophore system is essential for viability of Aspergillus
RT nidulans: functional analysis of two genes encoding l-ornithine N 5-
RT monooxygenase (sidA) and a non-ribosomal peptide synthetase (sidC).";
RL Mol. Microbiol. 49:359-375(2003).
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
DR EMBL; AV223812; AAP56239.1; -.
DR GO; GO:0003824; P: catalytic activity; IEA.
DR GO; GO:0048037; F: cofactor binding; IEA.
DR GO; GO:0008152; P: metabolism; IEA.
DR InterPro; IPR010071; AA: adenyldom.
DR InterPro; IPR001242; AMP-bind.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR006162; Peptidase.
DR InterPro; IPR006163; Pp_bind.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00668; Condensation; 4.
DR Pfam; PF00550; PP-binding; 5.
DR TIGRfam; TIGR01733; AA-adenyl-dom; 3.
DR PROSITE; PS00075; ACP DOMAIN; 5.
DR PROSITE; PS00455; AMP-BINDING; 3.
DR PROSITE; PS00697; DNA_LIGASE A1; UNKNOWN 1.
DR PROSITE; PS00012; PHOSPHOTRANSFERASE; UNKNOWN 1.
SQ SEQUENCE 4793 AA; 525432 MW; DD4186E991908601 CRC64;

Query Match 52.4%; Score 44; DB 2; Length 4793;
Best Local Similarity 60.0%; Pred. No. 1.2e+03;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 LAKSKRNPANLTPP 16
DB 1406 LAKSKANPSPITAP 1420
|:|||||:|:|:|

RESULT 14
Q74N70 PRELIMINARY; PRT; 167 AA.
AC Q74N70;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE NEQ071.
GN OrderedLocusNames=NEQ071;
OS Nanoarchaeum equitans.
OC Archaea; Nanoarchaeota; Nanoarchaeum.
OX NCBI_TaxID=160232;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Kin4-M;
RX MEDLINE=22946215; PubMed=14566062;
RA Waters E., Hohn M.J., Abel I., Graham D.E., Adams M.D., Barnstead M.,
RA Beeson K.Y., Bibbs L., Bolanos R., Keller M., Kretz K., Lin X.,
RA Mathur E., Ni J., Podar M., Richardson T., Sutton G.G., Simon M.,
RA Soell D., Stetter K.O., Short J.M., Noorderwier M.;
RT "The genome of Nanoarchaeum equitans: insights into early archaeal
RT evolution and derived parasitism.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:12984-12988(2003).
DR EMBL; AE017199; AAR38927.1; -.
KW Complete proteome.
SQ SEQUENCE 167 AA; 20126 MW; CCB6FD3D50EA7483 CRC64;

Query Match 51.2%; Score 43; DB 2; Length 167;
Best Local Similarity 64.3%; Pred. No. 45;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LAKSKRNPANLTPP 15
DB 141 LAKYKKNPIYLLP 154
|:|||||:|:|:|

RESULT 15
AAR38927 PRELIMINARY; PRT; 167 AA.
AC AAR38927;
DT 02-MAR-2004 (TRENBLrel. 27, Created)
DT 02-MAR-2004 (TRENBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TRENBLrel. 27, Last annotation update)
DE NEQ071.
GN NEQ071.
OS Nanoarchaeum equitans.
OC Archaea; Nanoarchaeota; Nanoarchaeum.
OX NCBI_TaxID=160232;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Kin4-M;
RX MEDLINE=22946215; PubMed=14566062;
RA Waters E., Hohn M.J., Abel I., Graham D.E., Adams M.D., Barnstead M.,
RA Beeson K.Y., Bibbs L., Bolanos R., Keller M., Kretz K., Lin X.,
RA Mathur E., Ni J., Podar M., Richardson T., Sutton G.G., Simon M.,
RA Soell D., Stetter K.O., Short J.M., Noorderwier M.;
RT "The genome of Nanoarchaeum equitans: insights into early archaeal
RT evolution and derived parasitism.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:12984-12988(2003).
DR EMBL; AE017199; AAR38927.1; -.
KW Complete proteome.
SQ SEQUENCE 167 AA; 20126 MW; CCB6FD3D50EA7483 CRC64;

Query Match 51.2%; Score 43; DB 2; Length 167;
Best Local Similarity 64.3%; Pred. No. 45;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LAKSKRNPANLTPP 15
DB 141 LAKYKKNPIYLLP 154
|:|||||:|:|:|

Search completed: November 10, 2004, 13:38:44

```

Job time : 22.5371 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 6.90674 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-26

Perfect score: 148

Sequence: 1 ANLLLMVPILIAMAFMLTERKILGYIQPR 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131	88.5	318	1 DNHUN1	NADH2 dehydrogenas
2	131	88.5	318	2 A59153	NADH2 dehydrogenas
3	117	79.1	318	2 T11833	NADH2 dehydrogenas
4	110	74.3	315	1 OXMS1M	NADH2 dehydrogenas
5	106	71.6	318	2 T11363	NADH2 dehydrogenas
6	106	71.6	318	2 T11857	NADH2 dehydrogenas
7	106	71.6	318	2 T11247	NADH2 dehydrogenas
8	105	70.9	318	2 T11480	NADH2 dehydrogenas
9	104	70.3	305	2 S52972	NADH2 dehydrogenas
10	104	70.3	318	1 OXBO1M	NADH2 dehydrogenas
11	104	70.3	318	2 A58888	NADH2 dehydrogenas
12	103	69.6	318	2 T11428	NADH2 dehydrogenas
13	100	67.6	318	2 S47870	NADH2 dehydrogenas
14	100	67.6	318	2 T10972	NADH2 dehydrogenas
15	100	67.6	323	1 OXLL1M	NADH2 dehydrogenas
16	99	66.9	318	1 QORT1M	NADH2 dehydrogenas
17	99	66.9	318	2 T11441	NADH2 dehydrogenas
18	99	66.9	318	2 T11410	NADH2 dehydrogenas
19	99	66.9	318	2 T11402	NADH2 dehydrogenas
20	99	66.9	318	2 A58850	NADH2 dehydrogenas
21	99	66.9	318	2 S41820	NADH2 dehydrogenas
22	99	66.9	318	2 T11050	NADH2 dehydrogenas
23	98	66.2	317	2 T11337	NADH2 dehydrogenas
24	98	66.2	318	2 T11493	NADH2 dehydrogenas
25	98	66.2	318	2 S41835	NADH2 dehydrogenas
26	98	66.2	318	2 S26151	NADH2 dehydrogenas
27	97	65.5	323	2 T11820	NADH2 dehydrogenas
28	96	64.9	313	2 B71390	NADH2 dehydrogenas
29	96	64.9	319	2 T11454	NADH2 dehydrogenas

30	96	64.9	322	2 S68128	NADH2 dehydrogenas
31	96	64.9	325	2 T11116	NADH2 dehydrogenas
32	94	63.5	334	2 T11890	NADH2 dehydrogenas
33	93	62.8	318	2 T11506	NADH2 dehydrogenas
34	92	62.2	322	2 A90620	NADH2 dehydrogenas
35	92	62.2	323	2 A34284	NADH2 dehydrogenas
36	92	62.2	323	2 A58892	NADH2 dehydrogenas
37	92	62.2	323	2 T10986	NADH2 dehydrogenas
38	91	61.5	318	2 T11389	NADH2 dehydrogenas
39	91	61.5	322	2 T11127	NADH2 dehydrogenas
40	91	61.5	323	2 T11793	NADH2 dehydrogenas
41	90	60.8	323	2 T09857	NADH2 dehydrogenas
42	90	60.8	324	2 T11534	NADH2 dehydrogenas
43	90	60.8	324	2 S36002	NADH2 dehydrogenas
44	90	60.8	324	2 S35462	NADH2 dehydrogenas
45	90	60.8	324	2 T09947	NADH2 dehydrogenas

ALIGNMENTS

RESULT 1

DNHUN1

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - human mitochondrion
N;Alternate names: NADH-ubiquinone oxidoreductase chain 1

C;Species: mitochondrion Homo sapiens (man)

C;Date: 22-May-1981 #sequence_revision 23-Oct-1981 #text_change 09-Jul-2004

C;Accession: A00407

R;Anderson, S.; Bankier, A.T.; Barrell, B.G.; de Bruijn, M.H.L.; Coulson, A.R.; Drouin, C.

Nature 290, 457-465, 1981

A;Title: Sequence and organization of the human mitochondrial genome.

A;Reference number: A00151; MUID:81173052; PMID:7219534

A;Accession: A00407

A;Molecule type: DNA

A;Residues: 1-318 <AND>

A;Cross-references: UNIPROT:P03886; GB:J01415; GB:M12548; GB:M58503; GB:M63932; GB:M63933;

C;Genetics:

A;Gene: GDB:MTND1

A;Cross-references: GDB:118911; OMIM:516000

A;Map position: MTH3307-4262

A;Genome: mitochondrion

A;Genetic code: SGCI

C;Superfamily: NADH:ubiquinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-t
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 88.5%; Score 131; DB 1; Length 318;

Best Local Similarity 90.3%; Pred. No. 1.2e-10;

Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ANLLLMVPILIAMAFMLTERKILGYIQPR 31

Db 4 ANLLLMVPILIAMAFMLTERKILGYMQLR 34

RESULT 2

A59153

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - western lowland gorilla mitochondrion
N;Alternate names: NADH-ubiquinone oxidoreductase chain 1

C;Species: mitochondrion Gorilla gorilla gorilla (western lowland gorilla)

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C;Accession: A59153

R;Xu, X.; Arnason, U.

Mol. Biol. Evol. 13, 691-698, 1996

A;Title: A complete sequence of the mitochondrial genome of the Western lowland gorilla.

A;Reference number: Z17269; MUID:96212991; PMID:8676744

A;Accession: A59153

A;Status: preliminary; nucleic acid sequence not shown; translation not shown; translated

A;Molecule type: DNA

A;Residues: 1-318 <XUX>

A;Cross-references: UNIPROT:Q9T9Z0; GB:X93347; NID:gl304307; GSPDB:GN00106

A;Note: Submitted to GenBank, November 1995

A;Note: this translation is not annotated in GenBank entry GGMITG, release 114.0

C:Genetics:

A:Gene: NADH1
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 C:Superfamily: NADH:quinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-t
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 88.5%; Score 131; DB 2; Length 318;

Best Local Similarity 90.3%; Pred. No. 1.2e-10;
 Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPIILAMAFMLTERKILGYIQPR 31

Db 4 ANLLLVPIILAMAFMLTERKILGYMQLR 34

RESULT 3

T11833

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - common gibbon mitochondrion
 C:Species: mitochondrion Hylobates lar (common gibbon, white-handed gibbon)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004

C:Accession: T11833

R:Aranson, U.; Gullberg, A.; Xu, X.

Hereditas 124, 185-189, 1996

A:Title: A complete mitochondrial DNA molecule of the white-handed gibbon, Hylobates lar

A:Reference number: Z17353

A:Accession: T11833

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-318 <ARN>

A:Cross-references: UNIPROT:Q96126; EMBL:X99256; PIDN:CAA67628.1

A:Experimental source: isolate Ester

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SGC1

A:Note: NADH1

C:Superfamily: NADH:quinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-t
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match

Best Local Similarity 79.1%; Score 117; DB 2; Length 318;

Matches 25; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 NLLLMVPIILAMAFMLTERKILGYIQPR 31

Db 5 NLLLLPILAMAFMLTERKILGYTQLR 34

RESULT 4

QXMS1M

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - mouse mitochondrion

A:Alternate names: NADH-ubiquinone oxidoreductase chain 1

C:Species: mitochondrion Mus musculus (house mouse)

C:Date: 02-Apr-1982 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004

C:Accession: A00409

R:Bibb, M.J.; Van Etten, R.A.; Wright, C.T.; Walberg, M.W.; Clayton, D.A.

Cell 26, 167-180, 1981

A:Title: Sequence and gene organization of mouse mitochondrion DNA.

A:Reference number: A00153; MUID:82137051; PMID:7332926

A:Accession: A00409

A:Molecule type: DNA

A:Residues: 1-315 <BIB>

A:Cross-references: UNIPROT:P03888; GB:J01420; NID:9342520; PIDN:AAB48644.1; PID:g896295

A:Note: the authors translated the initiation codon ATT for residue 1 as lle

C:Genetics:

A:Gene: NDI

A:Genome: mitochondrion

A:Genetic code: SGC1

A:Start codon: ATT

C:Superfamily: NADH:quinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-t

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match

74.3%; Score 110; DB 1; Length 315;

Best Local Similarity 76.7%; Pred. No. 9.4e-08;
 Matches 23; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 NLLLMVPIILAMAFMLTERKILGYIQPR 31

Db 2 NLLTLVPILAMAFMLTERKILGYMQLR 31

RESULT 5

T11363

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - donkey mitochondrion

C:Species: mitochondrion Equus asinus (donkey)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004

C:Accession: T11363

R:Xu, X.; Gullberg, A.; Aranson, U.

J. Mol. Evol. 43, 438-463, 1996

A:Title: The complete mitochondrial (mtDNA) of the donkey and mtDNA comparisons among for

A:Reference number: Z17265; MUID:97032591; PMID:8875857

A:Accession: T11363

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-318 <XUX>

A:Cross-references: UNIPROT:P92475; EMBL:X97337; NID:g1805746; PIDN:CAA66014.1; PID:g1805

A:Experimental source: kidney

C:Genetics:

A:Genome: mitochondrion

C:Superfamily: NADH:quinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-t

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; c

Query Match

Best Local Similarity 71.6%; Score 106; DB 2; Length 318;

Matches 21; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 NLLLMVPIILAMAFMLTERKILGYIQPR 31

Db 5 NVLLIIPILLAVAFMLTERKILGYMQLR 34

RESULT 6

T11857

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - horse mitochondrion

C:Species: mitochondrion Equus caballus (domestic horse)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004

C:Accession: T11857

R:Xu, X.; Aranson, U.

Gene 148, 357-362, 1994

A:Title: The complete mitochondrial DNA sequence of the horse, Equus caballus: Extensive

A:Reference number: Z17369; MUID:95047450; PMID:7958969

A:Accession: T11857

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-318 <XUX>

A:Cross-references: UNIPROT:P48652; EMBL:X79547; NID:g577571; PID:g577572; PIDN:CAA56079.

C:Genetics:

A:Genome: mitochondrion

C:Superfamily: NADH:quinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-t

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; c

Query Match

Best Local Similarity 71.6%; Score 106; DB 2; Length 318;

Matches 21; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 NLLLMVPIILAMAFMLTERKILGYIQPR 31

Db 5 NVLLIIPILLAVAFMLTERKILGYMQLR 34

RESULT 7

T11247

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - greater Indian rhinoceros mitoch

C:Species: mitochondrion Rhinoceros unicornis (greater Indian rhinoceros)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004

C:Accession: T11247

R;Xu, X.; Janke, A.; Arnason, U.
Mol. Biol. Evol. 13, 1167-1173, 1996
A;Title: The complete mitochondrial DNA sequence of the greater indian rhinoceros, Rhinoceros unicornis.
A;Reference number: Z17256; MUID:97051708; PMID:8896369
A;Accession: T11247
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-318 <XUD>
A;Cross-references: UNIPROT:Q96189; EMBL:X97336; NID:gl666193; PIDN:CAA66001.1; PID:gl666193
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC1
A;Superfamily: NADH:quinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-ty
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query Match 71.6%; Score 106; DB 2; Length 318;
Best Local Similarity 66.7%; Pred. No. 3.4e-07;
Matches 20; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
QY 2 NLLLMVPILIAMAFMLTERKILGYIQPR 31
Db 5 NLLLMVPILIAMAFMLTERKILGYIQPR 34
RESULT 8
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - rabbit mitochondrion (fragment)
C;Species: mitochondrion Oryctolagus cuniculus (domestic rabbit)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T11480
R;Gissi, C.; Gullberg, A.; Arnason, U.
Genomics 50, 161-169, 1998
A;Title: The complete mitochondrial DNA sequence of the rabbit, Oryctolagus cuniculus.
A;Reference number: Z17275; MUID:98317530; PMID:9653643
A;Accession: T11480
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-318 <GIS>
A;Cross-references: UNIPROT:O79427; EMBL:AJ001588; NID:g3293006; PIDN:CAA04847.1; PID:g3293006
C;Genetics:
A;Gene: NADH1
A;Genome: mitochondrion
A;Genetic code: SGC1
A;Superfamily: NADH:quinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-ty
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query Match 70.9%; Score 105; DB 2; Length 318;
Best Local Similarity 70.0%; Pred. No. 4.7e-07;
Matches 21; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
QY 2 NLLLMVPILIAMAFMLTERKILGYIQPR 31
Db 5 NLLLMVPILIAMAFMLTERKILGYIQPR 34
RESULT 9
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - honeybee mitochondrion
C;Species: mitochondrion Apis mellifera (honeybee)
C;Date: 15-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S52972
R;Crozier, R.H.; Crozier, Y.C.
Genetics 133, 97-117, 1993
A;Title: The mitochondrial genome of the honeybee Apis mellifera: complete sequence and
A;Reference number: S52960; MUID:93114603; PMID:8417993
A;Accession: S52972
A;Molecule type: DNA
A;Residues: 1-305 <CRO>
A;Cross-references: UNIPROT:P34847; EMBL:L06178; NID:g336279; PIDN:AAB96810.1; PID:g8290
A;Experimental source: ligustica
C;Genetics:

A;Genome: mitochondrion
A;Genetic code: SGC4
A;Start codon: GTG
C;Superfamily: NADH:quinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-ty
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query Match 70.3%; Score 104; DB 2; Length 305;
Best Local Similarity 70.0%; Pred. No. 6.2e-07;
Matches 21; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
QY 2 NLLLMVPILIAMAFMLTERKILGYIQPR 31
Db 6 NLLLMVPILIAMAFMLTERKILGYIQPR 35
RESULT 10
OXBOLM
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - bovine mitochondrion
N;Alternate names: NADH dehydrogenase (ubiquinone) 29k chain; NADH-ubiquinone oxidoreduct
C;Species: mitochondrion Bos primigenius taurus (cattle)
C;Date: 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change 09-Jul-2004
C;Accession: A00408; A31910
R;Anderson, S.; de Bruijn, M.H.L.; Coulson, A.R.; Eperon, I.C.; Sanger, F.; Young, I.G.
J. Mol. Biol. 156, 683-717, 1982
A;Title: Complete sequence of bovine mitochondrial DNA. Conserved features of the mammal
A;Reference number: A00152; MUID:83010260; PMID:7120390
A;Accession: A00408
A;Molecule type: DNA
A;Residues: 1-318 <AND>
A;Cross-references: UNIPROT:P03887; GB:J01394; NID:g336430; PIDN:AAB59268.1; PID:g336431;
R;Yagi, T.; Hatefi, Y.
J. Biol. Chem. 263, 16150-16155, 1988
A;Title: Identification of the dicyclohexylcarbodiimide-binding subunit of NADH-ubiquinol
A;Reference number: A31910; MUID:89034077; PMID:3141400
A;Accession: A31910
A;Molecule type: protein
A;Residues: 1-15 <YAG>
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC1
C;Superfamily: NADH:quinone oxidoreductase (complex I), subunit 1/[NiFe]-hydrogenase-3-ty
C;Keywords: blocked amino end; membrane-associated complex; mitochondrion; NAD; oxidative
F;1/Modified site: N-formylmethionine #status experimental
Query Match 70.3%; Score 104; DB 1; Length 318;
Best Local Similarity 63.3%; Pred. No. 6.4e-07;
Matches 19; Conservative 8; Mismatches 3; Indels 0; Gaps 0;
QY 2 NLLLMVPILIAMAFMLTERKILGYIQPR 31
Db 5 NLLLMVPILIAMAFMLTERKILGYIQPR 34
RESULT 11
A58888
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - duckbill platypus mitochondrion
C;Species: mitochondrion Ornithorhynchus anatinus (duckbill platypus)
C;Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 09-Jul-2004
C;Accession: A58888
R;Janke, A.; Gemmell, N.J.; Feldmaier-Fuchs, G.; von Haeseler, A.; Paabo, S.
J. Mol. Evol. 42, 153-159, 1996
A;Title: The mitochondrial genome of a monotreme--the platypus (Ornithorhynchus anatinus)
A;Reference number: A58888; MUID:97077300; PMID:8919867
A;Accession: A58888
A;Status: nucleic acid sequence not shown; translation not shown; not compared with conce
A;Molecule type: DNA
A;Residues: 1-318 <JAN>
A;Cross-references: UNIPROT:Q37717; GB:X83427; NID:g1469249; PIDN:CAA58455.1; PID:g14692;
A;Note: submitted to GenBank/EMBL/DBJ December, 1994
C;Genetics:
A;Map position: FOR2807-3763
A;Genome: mitochondrion
A;Genetic code: SGC1

1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 37.8532 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-26
Perfect score: 148
Sequence: 1 ANLLLMVPIILAMAFMLTERKILGYIQPR 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825:181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133	89.9	318	2 AAO66722	Aao66722 homo sapi
2	133	89.9	318	2 AAT37866	Aat37866 homo sapi
3	131	88.5	316	2 Q85KT1	Q85kt1 homo sapien
4	131	88.5	316	2 Q8WCX9	Q8wcx9 homo sapien
5	131	88.5	318	1 NUIM HUMAN	P03886 homo sapien
6	131	88.5	318	2 Q8R0U6	Q8rou6 homo sapien
7	131	88.5	318	2 Q8R0X2	Q8rox2 homo sapien
8	131	88.5	318	2 Q8RR12	Q8rr12 homo sapien
9	131	88.5	318	2 Q8RR39	Q8rr39 homo sapien
10	131	88.5	318	2 Q8RR93	Q8rr93 homo sapien
11	131	88.5	318	2 Q8RRM1	Q8rrm1 homo sapien
12	131	88.5	318	2 Q8VH72	Q8vh72 homo sapien
13	131	88.5	318	2 Q8VIF0	Q8vifo homo sapien
14	131	88.5	318	2 Q8WQ94	Q8wq94 homo sapien
15	131	88.5	318	2 Q7YU22	Q7yu22 homo sapien
16	131	88.5	318	2 Q7YCD0	Q7ycd0 homo sapien
17	131	88.5	318	2 Q7YCD4	Q7ycd4 homo sapien
18	131	88.5	318	2 Q7YCF4	Q7ycf4 homo sapien
19	131	88.5	318	2 Q7YCG9	Q7ycg9 homo sapien
20	131	88.5	318	2 Q7YED6	Q7yed6 homo sapien
21	131	88.5	318	2 Q7YEG1	Q7yeg1 homo sapien
22	131	88.5	318	2 Q85BK7	Q85bk7 homo sapien
23	131	88.5	318	2 Q85KT8	Q85kt8 homo sapien
24	131	88.5	318	2 Q85KU3	Q85ku3 homo sapien
25	131	88.5	318	2 Q85KU5	Q85ku5 homo sapien
26	131	88.5	318	2 Q85KW9	Q85kw9 homo sapien
27	131	88.5	318	2 Q85KX9	Q85kx9 homo sapien
28	131	88.5	318	2 Q85L02	Q85l02 homo sapien
29	131	88.5	318	2 Q85L07	Q85l07 homo sapien
30	131	88.5	318	2 Q8HHH1	Q8hhh1 homo sapien
31	131	88.5	318	2 Q8HHH2	Q8hhh2 homo sapien

```

32 131 88.5 318 2 Q8HHH3      Q8hhh3 homo sapien
33 131 88.5 318 2 Q9T9Z0      Q9t9z0 gorilla gor
34 131 88.5 318 2 Q9B1L5      Q9b1l5 homo sapien
35 131 88.5 318 2 Q9B1M4      Q9b1m4 homo sapien
36 131 88.5 318 2 Q9B2U3      Q9b2u3 homo sapien
37 131 88.5 318 2 Q9B2V7      Q9b2v7 homo sapien
38 131 88.5 318 2 Q9B2X4      Q9b2x4 homo sapien
39 131 88.5 318 2 Q9B2X5      Q9b2x5 homo sapien
40 131 88.5 318 2 Q8WCX2      Q8wcx2 homo sapien
41 131 88.5 318 2 AAL54384      Aal54384 homo sapi
42 131 88.5 318 2 AAL54410      Aal54410 homo sapi
43 131 88.5 318 2 AAL54423      Aal54423 homo sapi
44 131 88.5 318 2 AAL54449      Aal54449 homo sapi
45 131 88.5 318 2 AAL54475      Aal54475 homo sapi

```

ALIGNMENTS

```

RESULT 1
AAO66722
ID AAO66722 PRELIMINARY; PRT; 318 AA.
AC AAO66722;
DT 24-MAY-2004 (Tremblrel. 27, Created)
DT 24-MAY-2004 (Tremblrel. 27, Last sequence update)
DT 24-MAY-2004 (Tremblrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 1.
GN NDI.
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22807936; PubMed=12870132;
RA Kong Q.-P., Yao Y.-G., Sun C., Bandelt H.-J., Zhu C.-L., Zhang Y.-P.;
RT "Phylogeny of east Asian mitochondrial DNA lineages inferred from
RT complete sequences."
RL Am. J. Hum. Genet. 73:671-676(2003).
DR EMBL; AY255141; AAO66722.1; -.
KW Mitochondrion.
SQ SEQUENCE 318 AA; 35628 MW; 8B265AD2ECE7B297 CRC64;

```

```

Query Match 89.9%; Score 133; DB 2; Length 318;
Best Local Similarity 90.3%; Pred. No. 1.9e-09;
Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 ANLLLMVPIILAMAFMLTERKILGYIQPR 31
      |||||:|||||:|||||:|||||:|
DB 4 ANLLLMVPIILAMAFMLTERKILGYVQLR 34

```

```

RESULT 2
AAT37866
ID AAT37866 PRELIMINARY; PRT; 318 AA.
AC AAT37866;
DT 01-JUN-2004 (Tremblrel. 27, Created)
DT 01-JUN-2004 (Tremblrel. 27, Last sequence update)
DT 01-JUN-2004 (Tremblrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 1.
GN NDI.
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NgasanM1030HGc;
RA Starikovskaya E.B., Sukernik R.I., Derbenev O.A., Volodko N.V.,
RA Ruiz-Pesini E., Torroni A., Brown M.D., Lott M.T., Hosseini S.H.,

```

RA Huoponen K., Wallace D.C.;
 RT "Mitochondrial DNA diversity in indigenous populations of southern
 RL extent Siberia, and the origins of native american haplogroups.";
 RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY615359; AAT37866.1; -
 KW Mitochondrion.
 SQ SEQUENCE 318 AA; 35628 MW; 8B265AD2ECE7B297 CRC64;

Query Match 89.9%; Score 133; DB 2; Length 316;
 Best Local Similarity 90.3%; Pred. No. 1.9e-09;
 Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPIILAMAFMLTERKILGYIQPR 31
 |||||:|||||:|||||:|||||:|||||:
 Db 4 ANLLLMVPIILAMAFMLTERKILGYQLR 34

RESULT 3

Q85KT1 PRELIMINARY; PRT; 316 AA.
 AC Q85KT1;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE NADH dehydrogenase subunit 1.
 GN Name=ND1;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=22406325; PubMed=12509511;
 RA Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G.,
 RA Hosseini S., Brandon M., Easley K., Chen E., Brown M.D.,
 RA Sukernik R.I., Olickers A., Wallace D.C.;
 RT "Natural selection shaped regional mtDNA variation in humans.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:171-176 (2003).
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -!- SIMILARITY: Belongs to the complex I subunit 1 family.
 DR EMBL; AY195783; AAO88774.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001694; Resp_NADH_dhl.
 DR Pfam; PF00146; NADHdh; 1.
 DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
 DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
 KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
 SQ SEQUENCE 316 AA; 35447 MW; 3A64BB008E2DE84E CRC64;

Query Match 88.5%; Score 131; DB 2; Length 316;
 Best Local Similarity 90.3%; Pred. No. 3.6e-09;
 Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPIILAMAFMLTERKILGYIQPR 31
 |||||:|||||:|||||:|||||:|||||:
 Db 2 ANLLLMVPIILAMAFMLTERKILGYQLR 32

RESULT 4

Q8WCX9 PRELIMINARY; PRT; 316 AA.
 AC Q8WCX9;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE NADH dehydrogenase subunit 1.
 GN Name=ND1;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=11593319;
 RA Maca-Meyer N., Gonzalez A.M., Larruga J.M., Flores C., Cabrera V.M.;
 RT "Major genomic mitochondrial lineages delineate early human
 RL expansions.";
 RL BMC Genet. 2:13-13 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cabrera V.M., Maca-Meyer N., Gonzalez A.M., Larruga J.M., Flores C.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -!- SIMILARITY: Belongs to the complex I subunit 1 family.
 DR EMBL; AF381994; AAL54553.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001694; Resp_NADH_dhl.
 DR Pfam; PF00146; NADHdh; 1.
 DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
 DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
 KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
 SQ SEQUENCE 316 AA; 35432 MW; 025E88C09BF7D7BD CRC64;

Query Match 88.5%; Score 131; DB 2; Length 316;
 Best Local Similarity 90.3%; Pred. No. 3.6e-09;
 Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPIILAMAFMLTERKILGYIQPR 31
 |||||:|||||:|||||:|||||:|||||:
 Db 2 ANLLLMVPIILAMAFMLTERKILGYQLR 32

RESULT 5

NUIM HUMAN STANDARD; PRT; 318 AA.
 ID NUIM HUMAN
 AC P03856; Q37523;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 1 (BC 1.6.5.3).
 GN Name=MTND1; Synonyms=ND1;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81173052; PubMed=7219534;
 RA Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L.,
 RA Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A.,
 RA Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G.;
 RT "Sequence and organization of the human mitochondrial genome.";
 RL Nature 290:457-465 (1981).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANTS ALA-87 AND ALA-168.
 RC TISSUE=Placenta;
 RX MEDLINE=9532634; PubMed=7530363;
 RA Horai S., Hayasaka K., Kondo R., Tsugane K., Takahata N.;
 RT "Recent African origin of modern humans revealed by complete sequences
 RL of hominoid mitochondrial DNAs.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:532-536 (1995).
 RN [3]
 RP SEQUENCE FROM N.A.

RX PubMed=12949126; DOI=10.1093/molbev/msg230;
 RA Moilanen J.S., Finnila S., Majamaa K.;
 RT "Lineage-specific selection in human mtDNA: lack of polymorphisms in a
 RL segment of MTND5 gene in haplogroup J.";
 RL Mol. Biol. Evol. 20:2132-2142 (2003).
 RN [4]

RP SEQUENCE FROM N.A.
RX MEDLINE=21012010; PubMed=11130070; DOI=10.1038/35047064;
RA Ingman M., Kaessmann H., Paabo S., Gyllenstein U.;
RT "Mitochondrial genome variation and the origin of modern humans.";
RL Nature 408:708-713(2000).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=22723755; PubMed=12840039; DOI=10.1101/gr.686603;
RA Ingman M., Gyllenstein U.;
RT "Mitochondrial genome variation and evolutionary history of Australian
and new guinean aborigines.";
RL Genome Res. 13:1600-1606(2003).
RN [6]
RP SEQUENCE FROM N.A.
RX PubMed=14760490; DOI=10.1007/s00414-004-0427-6;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 118:137-146(2004).
RN [7]
RP SEQUENCE OF 130-318 FROM N.A.
RX MEDLINE=81170577; PubMed=6260957;
RA Sanger F., Coulson A.R., Barrell B.G., Smith A.J.H., Roe B.A.;
RT "Cloning in single-stranded bacteriophage as an aid to rapid DNA
sequencing.";
RL J. Mol. Biol. 143:161-178(1980).
RN [8]
RP IDENTIFICATION OF PROTEIN.
RX MEDLINE=85188293; PubMed=3921850;
RA Chomay A., Mariotti P., Clester M.W.J., Ragan C.I., Matsuno-Yagi A.,
RA Hattefi Y., Doellittle R.F., Attardi G.;
RT "Six unidentified reading frames of human mitochondrial DNA encode
components of the respiratory-chain NADH dehydrogenase.";
RL Nature 314:592-597(1985).
RN [9]
RP VARIANT LHON THR-52.
RX MEDLINE=92026096; PubMed=1928099;
RA Howell N., Bindoff L.A., McCullough D.A., Kubacka I., Poulton J.,
RA Mackey D., Taylor L., Turnbull D.M.;
RT "Leber hereditary optic neuropathy: identification of the same
mitochondrial ND1 mutation in six pedigrees.";
RL Am. J. Hum. Genet. 49:939-950(1991).
RN [10]
RP VARIANT LHON THR-52.
RX MEDLINE=91241131; PubMed=1674640;
RA Huoponen K., Vilkkii J., Aula P., Nikoskelainen E.K., Savontaus M.L.;
RT "A new mtDNA mutation associated with Leber hereditary optic
neuropathy.";
RL Am. J. Hum. Genet. 48:1147-1153(1991).
RN [11]
RP VARIANT LHON PRO-285, AND VARIANT CYS-277.
RX MEDLINE=91206406; PubMed=2018041;
RA Howell N., Kubacka I., Xu M., McCullough D.A.;
RT "Leber hereditary optic neuropathy: involvement of the mitochondrial
ND1 gene and evidence for an intragenic suppressor mutation.";
RL Am. J. Hum. Genet. 48:935-942(1991).
RN [12]
RP VARIANT LHON HIS-304.
RX MEDLINE=91144615; PubMed=1900003;
RA Johns D.R., Berman J.;
RT "Alternative simultaneous complex I mitochondrial DNA mutations in
Leber's hereditary optic neuropathy.";
RL Biochem. Biophys. Res. Commun. 174:1324-1330(1991).
RN [13]
RP CHARACTERIZATION OF VARIANT LHON THR-52.
RX MEDLINE=92070510; PubMed=1959619;
RA Majander A., Huoponen K., Savontaus M.-L., Nikoskelainen E.,
RA Wikstrom M.;
RT "Electron transfer properties of NADH:ubiquinone reductase in the
ND1/3460 and the ND4/11778 mutations of the Leber hereditary optic
neuropathy (LHON).";
RL FEBS Lett. 292:289-292(1991).
RN [14]
RP VARIANTS PRO-205; CYS-255 AND PRO-288.
RX MEDLINE=92098084; PubMed=1757091;
RA Marzuki S., Noer A.S., Lertrit P., Thyagarajan D., Kapsa R.,
RA Utthanaphol P., Byrne E.;
RT "Normal variants of human mitochondrial DNA and translation products:
the building of a reference data base.";
RL Hum. Genet. 88:139-145(1991).
RN [15]
RP VARIANT LHON HIS-30.
RX MEDLINE=93038635; PubMed=1417830;
RA Johns D.R., Neufeld M.J., Park R.D.;
RT "An ND-6 mitochondrial DNA mutation associated with Leber hereditary
optic neuropathy.";
RL Biochem. Biophys. Res. Commun. 187:1551-1557(1992).
RN [16]
RP VARIANT MELAS THR-31.
RX MEDLINE=94010883; PubMed=8104867;
RA Shoffner J.M., Brown M.D., Tortoni A., Lott M.T., Cabell M.F.,
RA Mirra S.S., Beal M.F., Yang C.-C., Gearing M., Salvo R., Watts R.L.,
RA Juncos J.L., Hansen L.A., Crain B.J., Payad M., Reckord C.L.,
RA Wallace D.C.;
RT "Mitochondrial DNA variants observed in Alzheimer disease and
Parkinson disease patients.";
RL Genomics 17:171-184(1993).
RN [17]
RP VARIANT MELAS THR-31.
RX MEDLINE=96303708; PubMed=8723687;
RA Jakisch M., Hofmann S., Kaufhold P., Obermaier-Kusser B., Zierz S.,
RA Gerbitz K.-D.;
RT "A novel combination of mitochondrial tRNA and ND1 gene mutations in a
syndrome with MELAS, cardiomyopathy, and diabetes mellitus.";
RL Hum. Mutat. 7:358-360(1996).
RN [18]
RP VARIANT NIDDM THR-4.
RX MEDLINE=95251666; PubMed=7733935;
RA Nakagawa Y., Ikegami H., Yamato E., Takekawa K., Fujisawa T.,
RA Hamada Y., Ueda H., Uchigata Y., Miki T., Kumahara Y.;
RT "A new mitochondrial DNA mutation associated with non-insulin-
dependent diabetes mellitus.";
RL Biochem. Biophys. Res. Commun. 209:664-668(1995).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- DISEASE: Defects in MTND1 are a cause of Leber's hereditary optic
neuropathy (LHON) [MIM:535000]; also known as Leber's optic
atrophy. LHON is a maternally inherited disease resulting in acute
bilateral blindness due to retinal degeneration predominantly in
young men. Cardiac conduction defects and neurological defects
have also been described, resulting in optic nerve degeneration
and cardiac dysrhythmia.
CC -!- DISEASE: Defects in MTND1 are a cause of mitochondrial
encephalomyopathy with lactic acidosis and stroke-like episodes
syndrome (MELAS) [MIM:540000]. MELAS is a genetically
heterogeneous disorder, characterized by episodic vomiting,
seizures, and recurrent cerebral insults resembling strokes and
causing hemiparesis, hemianopsia, or cortical blindness.
CC -!- DISEASE: Defects in MTND1 could be associated to mitochondrial
susceptibility to Alzheimer's disease (AD) [MIM:502500].
CC -!- DISEASE: Defects in MTND1 could be associated with non-insulin-
dependent diabetes mellitus (NIDDM).
CC -!- SIMILARITY: Belongs to the complex I subunit 1 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V00662; CAA24026.1; -.
CC EMBL; J01415; AAB58943.1; -.
CC EMBL; D38112; BAA07290.1; -.
CC EMBL; AY339402; AAF89036.1; -.

DR EMBL; AY339403; AAP89049.1; -
 DR EMBL; AY339404; AAP89062.1; -
 DR EMBL; AY339405; AAP89075.1; -
 DR EMBL; AY339406; AAP89088.1; -
 DR EMBL; AY339407; AAP89101.1; -
 DR EMBL; AY339408; AAP89114.1; -
 DR EMBL; AY339409; AAP89127.1; -
 DR EMBL; AY339410; AAP89140.1; -
 DR EMBL; AY339411; AAP89153.1; -
 DR EMBL; AY339412; AAP89166.1; -
 DR EMBL; AY339413; AAP89179.1; -
 DR EMBL; AY339414; AAP89192.1; -
 DR EMBL; AY339415; AAP89205.1; -
 DR EMBL; AY339416; AAP89218.1; -
 DR EMBL; AY339417; AAP89231.1; -
 DR EMBL; AY339418; AAP89244.1; -
 DR EMBL; AY339419; AAP89257.1; -
 DR EMBL; AY339420; AAP89270.1; -

Query Match 88.5%; Score 131; DB 1; Length 318;
 Best Local Similarity 90.3%; Pred. No. 3.6e-09;
 Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPILIAMAFMLTERKILGYIQPR 31
 |||||:|||||:|||||:|||||:|||||:
 DB 4 ANLLLMVPILIAMAFMLTERKILGYMQLR 34

RESULT 6

Q6R0U6 PRELIMINARY; PRT; 318 AA.

AC Q6R0U6; (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DE NADH dehydrogenase subunit 1.
 GN Name=ND1;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Starikovskaya E.B., Sukernik R.I., Derbeneva O.A., Volodko N.V.,
 RA Torroni A., Ruiz-Pesini E., Brown M.D., Lott M.T., Hosseini S.H.,
 RA Huoponen K., Wallace D.C.,
 RT "Mitochondrial DNA Diversity in Indigenous Populations of Southern
 Extent of Siberia, and the Origins of Native American Haplogroups.";
 RL Ann. Hum. Genet. 0:0-0(2004).
 CC -|- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -|- SIMILARITY: Belongs to the complex I subunit 1 family.
 DR EMBL; AY519490; AAR91284.2; -
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR001694; Resp_NADH_dhl.
 DR Pfam; PF00146; NADHdh; 1.
 DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
 DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
 DR Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
 KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
 SQ SEQUENCE 318 AA; 35669 MW; 2F7BD6C75CC295D CRC64;

Query Match 88.5%; Score 131; DB 2; Length 318;
 Best Local Similarity 90.3%; Pred. No. 3.6e-09;
 Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPILIAMAFMLTERKILGYIQPR 31
 |||||:|||||:|||||:|||||:|||||:
 DB 4 ANLLLMVPILIAMAFMLTERKILGYMQLR 34

RESULT 7

Q6R0X2 PRELIMINARY; PRT; 318 AA.

AC Q6R0X2; (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE NADH dehydrogenase subunit 1.
 GN Name=ND1;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Starikovskaya E.B., Sukernik R.I., Derbeneva O.A., Volodko N.V.,
 RA Torroni A., Ruiz-Pesini E., Brown M.D., Lott M.T., Hosseini S.H.,
 RA Huoponen K., Wallace D.C.,
 RT "Mitochondrial DNA Diversity in Indigenous Populations of Southern
 Extent of Siberia, and the Origins of Native American Haplogroups.";
 RL Ann. Hum. Genet. 0:0-0(2004).
 CC -|- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -|- SIMILARITY: Belongs to the complex I subunit 1 family.
 DR EMBL; AY519488; AAR91258.1; -
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR001694; Resp_NADH_dhl.
 DR Pfam; PF00146; NADHdh; 1.
 DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
 DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
 DR Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
 KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
 SQ SEQUENCE 318 AA; 35650 MW; 934495CA8BD0EA91 CRC64;

Query Match 88.5%; Score 131; DB 2; Length 318;
 Best Local Similarity 90.3%; Pred. No. 3.6e-09;
 Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPILIAMAFMLTERKILGYIQPR 31
 |||||:|||||:|||||:|||||:|||||:
 DB 4 ANLLLMVPILIAMAFMLTERKILGYMQLR 34

RESULT 8

Q6RR12 PRELIMINARY; PRT; 318 AA.

AC Q6RR12; (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE NADH dehydrogenase subunit 1.
 GN Name=ND1;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Starikovskaya E.B., Sukernik R.I., Derbeneva O.A., Volodko N.V.,
 RA Torroni A., Ruiz-Pesini E., Brown M.D., Lott M.T., Hosseini S.H.,
 RA Huoponen K., Wallace D.C.,
 RT "Mitochondrial DNA Diversity in Indigenous Populations of Southern
 Extent of Siberia, and the Origins of Native American Haplogroups.";
 RL Ann. Hum. Genet. 0:0-0(2004).
 CC -|- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -|- SIMILARITY: Belongs to the complex I subunit 1 family.
 DR EMBL; AY495134; AAR93068.1; -
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR001694; Resp_NADH_dhl.
 DR Pfam; PF00146; NADHdh; 1.
 DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
 DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
 DR Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
 KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
 SQ SEQUENCE 318 AA; 35630 MW; 2F721BCB95095591 CRC64;

Query Match 88.5%; Score 131; DB 2; Length 318;
 Best Local Similarity 90.3%; Pred. No. 3.6e-09;
 Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPILIAMAFMLTERKILGYIQPR 31
 |||||:|||||:|||||:|||||:|||||:
 DB 4 ANLLLMVPILIAMAFMLTERKILGYMQLR 34

RESULT 9

Q6R0X2 PRELIMINARY; PRT; 318 AA.

Best Local Similarity 90.3%; Pred. No. 3.6e-09;		Pred. No. 3.6e-09;		Pred. No. 3.6e-09;					
Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;		Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;		Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;					
1 ANLLLMVPIIIAMAFMLTERKILGYIQPR 31									
4 ANLLLVPIIIAMAFMLTERKILGYMQLR 34									
RESULT 9									
Q6RR39 PRELIMINARY; PRT; 318 AA.									
AC Q6RR39 (TREMBlrel. 27, Created)									
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)									
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)									
DE NADH dehydrogenase subunit 1.									
GN Name=NDL;									
OS Homo sapiens (Human).									
OG Mitochondrion.									
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
OX NCBI_TaxID=9606;									
RN [1]									
RP SEQUENCE FROM N.A.									
RX PubMed=14760490;									
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,									
RA Irwin J.A., Parsons T.J.;									
RT "Single nucleotide polymorphisms over the entire mtDNA genome that									
RT increase the power of forensic testing in Caucasians.";									
RL Int. J. Legal Med. 0:0-0(2004).									
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.									
CC -!- SIMILARITY: Belongs to the complex I subunit 1 family.									
DR EMBL; AY495131; AAR93029.1; -									
DR GO; GO:0005739; C:mitochondrion; IEA.									
DR InterPro; IPR001694; Resp_NADH_dhl.									
DR Pfam; PF00146; NADHdh; 1.									
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.									
DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.									
KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.									
RL Int. J. Legal Med. 0:0-0(2004).									
SQ SEQUENCE 318 AA; 35690 MW; D483B1C561AB2F31 CRC64;									
Query Match 88.5%; Score 131; DB 2; Length 318;									
Best Local Similarity 90.3%; Pred. No. 3.6e-09;									
Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;									
1 ANLLLMVPIIIAMAFMLTERKILGYIQPR 31									
4 ANLLLVPIIIAMAFMLTERKILGYMQLR 34									
RESULT 10									
Q6RR93 PRELIMINARY; PRT; 318 AA.									
AC Q6RR93 (TREMBlrel. 27, Created)									
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)									
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)									
DE NADH dehydrogenase subunit 1.									
GN Name=NDL;									
OS Homo sapiens (Human).									
OG Mitochondrion.									
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
OX NCBI_TaxID=9606;									
RN [1]									
RP SEQUENCE FROM N.A.									
RX PubMed=14760490;									
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,									
RA Irwin J.A., Parsons T.J.;									
RT "Single nucleotide polymorphisms over the entire mtDNA genome that									
RT increase the power of forensic testing in Caucasians.";									
RL Int. J. Legal Med. 0:0-0(2004).									
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.									
CC -!- SIMILARITY: Belongs to the complex I subunit 1 family.									
DR EMBL; AY495111; AAR92769.1; -									
DR GO; GO:0005739; C:mitochondrion; IEA.									
DR InterPro; IPR001694; Resp_NADH_dhl.									
DR Pfam; PF00146; NADHdh; 1.									
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.									
DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.									
KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.									
RL Int. J. Legal Med. 0:0-0(2004).									
SQ SEQUENCE 318 AA; 35690 MW; D483B1C561AB2F31 CRC64;									
Query Match 88.5%; Score 131; DB 2; Length 318;									
Best Local Similarity 90.3%; Pred. No. 3.6e-09;									
Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;									
1 ANLLLMVPIIIAMAFMLTERKILGYIQPR 31									
4 ANLLLVPIIIAMAFMLTERKILGYMQLR 34									
RESULT 11									
Q6RRM1 PRELIMINARY; PRT; 318 AA.									
AC Q6RRM1 (TREMBlrel. 27, Created)									
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)									
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)									
DE NADH dehydrogenase subunit 1.									
GN Name=NDI;									
OS Homo sapiens (Human).									
OG Mitochondrion.									
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
OX NCBI_TaxID=9606;									
RN [1]									
RP SEQUENCE FROM N.A.									
RX PubMed=14760490;									
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,									
RA Irwin J.A., Parsons T.J.;									
RT "Single nucleotide polymorphisms over the entire mtDNA genome that									
RT increase the power of forensic testing in Caucasians.";									
RL Int. J. Legal Med. 0:0-0(2004).									
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.									
CC -!- SIMILARITY: Belongs to the complex I subunit 1 family.									
DR EMBL; AY495111; AAR92769.1; -									
DR GO; GO:0005739; C:mitochondrion; IEA.									
DR InterPro; IPR001694; Resp_NADH_dhl.									
DR Pfam; PF00146; NADHdh; 1.									
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.									
DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.									
KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.									
RL Int. J. Legal Med. 0:0-0(2004).									
SQ SEQUENCE 318 AA; 35690 MW; 2EC48962B4892D09 CRC64;									
Query Match 88.5%; Score 131; DB 2; Length 318;									
Best Local Similarity 90.3%; Pred. No. 3.6e-09;									
Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;									
1 ANLLLMVPIIIAMAFMLTERKILGYIQPR 31									
4 ANLLLVPIIIAMAFMLTERKILGYMQLR 34									
RESULT 12									
Q6VH72 PRELIMINARY; PRT; 318 AA.									
ID Q6VH72									

```

AC Q6VH72;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 1.
GN Name=ND1;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12949126;
RA Moilanen J.S., Finnila S., Majamaa K.;
RT "Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J.";
RL Mol. Biol. Evol. 20:2132-2142(2003).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed=12949126;
RA Moilanen J.S., Finnila S., Majamaa K.;
RT "Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J.";
RL Mol. Biol. Evol. 20:2132-2142(2003).
RN [3]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T., Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -|- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -|- SIMILARITY: Belongs to the complex I subunit 1 family.
DR EMBL; AY339589; AAP91467.1; -
DR EMBL; AY339590; AAP91480.1; -
DR EMBL; AY495196; AAR93874.1; -
DR EMBL; AY495209; AAR94043.1; -
DR EMBL; AY339587; AAP91441.1; -
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR001694; Resp_NADH_dhl.
DR Pfam; PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
SQ SEQUENCE 318 AA; 35664 MW; 0895BF8C65002FA6 CRC64;

Query Match 88.5%; Score 131; DB 2; Length 318;
Best Local Similarity 90.3%; Pred. No. 3.6e-09;
Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPILIAMAFMLTKILGYIQPR 31
Db 4 ANLLLMVPILIAMAFMLTKILGYMQLR 34

RESULT 13
Q6VIF0 PRELIMINARY; PRT; 318 AA.
AC Q6VIF0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 1.
GN Name=ND1;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14563219;
RA Maca-Meyer N., Gonzalez A.M., Pestano J., Flores C., Larruga J.M., Cabrera V.M.;
RT "Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography.";
RL BMC Genet. 4:15-15(2003).
CC -|- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -|- SIMILARITY: Belongs to the complex I subunit 1 family.
DR EMBL; AY275534; AAQ19421.1; -
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR001694; Resp_NADH_dhl.
DR Pfam; PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
SQ SEQUENCE 318 AA; 35675 MW; 2F771ACB951CF45C CRC64;

Query Match 88.5%; Score 131; DB 2; Length 318;
Best Local Similarity 90.3%; Pred. No. 3.6e-09;
Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPILIAMAFMLTKILGYIQPR 31
Db 4 ANLLLMVPILIAMAFMLTKILGYMQLR 34

RESULT 15
Q6WQ94 PRELIMINARY; PRT; 318 AA.
AC Q6WQ94;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 1.
GN Name=ND1;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14563219;
RA Maca-Meyer N., Gonzalez A.M., Pestano J., Flores C., Larruga J.M., Cabrera V.M.;
RT "Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography.";
RL BMC Genet. 4:15-15(2003).
CC -|- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -|- SIMILARITY: Belongs to the complex I subunit 1 family.
DR EMBL; AY275534; AAQ19421.1; -
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR001694; Resp_NADH_dhl.
DR Pfam; PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
SQ SEQUENCE 318 AA; 35675 MW; 2F771ACB951CF45C CRC64;

Query Match 88.5%; Score 131; DB 2; Length 318;
Best Local Similarity 90.3%; Pred. No. 3.6e-09;
Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANLLLMVPILIAMAFMLTKILGYIQPR 31
Db 4 ANLLLMVPILIAMAFMLTKILGYMQLR 34

RESULT 15

```

```

Q7Y7U2
ID Q7Y7U2 PRELIMINARY; PRT; 318 AA.
AC Q7Y7U2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 1.
GN Name=ND1;
OS Homo sapiens (Human).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kong Q.-P., Yao Y.-G., Sun C., Bandelt H.-J., Zhu C.-L., Zhang Y.-P.;
RT "Phylogeny of East Asian Mitochondrial DNA Lineages Inferred from
RT Complete Sequences.";
RL Am. J. Hum. Genet. 0:0-0(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=22723755; PubMed=12840039;
RT Ingman M., Gyllenstein U.;
RT "Mitochondrial genome variation and evolutionary history of Australian
RT and New Guinean aborigines.";
RL Genome Res. 13:1600-1606(2003).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SIMILARITY: Belongs to the complex I subunit 1 family.
DR EMBL; AY255178; AAC67202.1; -.
DR EMBL; AY289062; AAP48023.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001694; Resp_NADH_dhl.
DR Pfam; PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_NDI_1; 1.
DR PROSITE; PS00668; COMPLEX1_NDI_2; 1.
KW Mitochondrion; NAD; Oxidoreductase; Transmembrane; Ubiquinone.
SQ SEQUENCE 318 AA; 35633 MW; 2F77061A58002591 CRC64;

Query Match 88.5%; Score 131; DB 2; Length 318;
Best Local Similarity 90.3%; Freq. No. 3.6e-09;
Matches 28; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ANLLLLMVPILIAAFMLMLTERKILGYIQPR 31
| | | | | | | | | | | | | | | | | |
Db 4 ANLLLLIVPILIAAFMLMLTERKILGYWQLR 34

Search completed: November 10, 2004, 13:38:23
Job time : 38.8532 secs

```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 10.3627 Seconds/sec
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-27

Perfect score: 150

Sequence: 1 LRLNTTWPITPILITLFLITNRLITR 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Issued Patents AA.*
- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	102.5	68.3	68	3	US-09-097-889-17
2	102.5	68.3	68	4	US-09-098-079-17
3	54	36.0	440	4	US-09-583-110-4693
4	52	34.7	107	4	US-09-248-796A-21089
5	52	34.7	176	4	US-09-248-796A-25834
6	52	34.7	286	4	US-09-138-452A-547
7	51.5	34.3	263	4	US-09-270-767-57472
8	51.5	34.3	456	4	US-09-270-767-42198
9	51	34.0	60	4	US-09-248-796A-26751
10	51	34.0	302	4	US-09-543-681A-8270
11	50	33.3	129	4	US-09-248-796A-20246
12	49	32.7	65	1	US-08-085-917-14
13	49	32.7	65	1	US-08-085-917-16
14	49	32.7	65	1	US-08-085-917-18
15	49	32.7	65	1	US-08-095-068-14
16	49	32.7	65	1	US-08-095-068-16
17	49	32.7	65	1	US-08-095-068-18
18	49	32.7	65	1	US-08-140-721A-14
19	49	32.7	65	1	US-08-140-721A-16
20	49	32.7	65	1	US-08-140-721A-18
21	49	32.7	65	1	US-08-619-790C-14
22	49	32.7	65	1	US-08-619-790C-16
23	49	32.7	65	1	US-08-619-790C-18
24	49	32.7	65	2	US-07-785-565A-14
25	49	32.7	65	2	US-07-785-565A-16
26	49	32.7	65	2	US-07-785-565A-18
27	49	32.7	71	1	US-08-085-917-15

28	49	32.7	71	1	US-08-055-917-17	Sequence 17, Appl
29	49	32.7	71	1	US-08-095-068-15	Sequence 15, Appl
30	49	32.7	71	1	US-08-095-068-17	Sequence 17, Appl
31	49	32.7	71	1	US-08-140-721A-15	Sequence 15, Appl
32	49	32.7	71	1	US-08-140-721A-17	Sequence 17, Appl
33	49	32.7	71	1	US-08-619-790C-15	Sequence 15, Appl
34	49	32.7	71	1	US-08-619-790C-17	Sequence 17, Appl
35	49	32.7	71	2	US-07-785-565A-15	Sequence 15, Appl
36	49	32.7	71	2	US-07-785-565A-17	Sequence 17, Appl
37	49	32.7	226	3	US-09-097-889-18	Sequence 18, Appl
38	49	32.7	226	4	US-09-098-079-18	Sequence 18, Appl
39	48	32.0	259	4	US-09-477-135A-133	Sequence 133, Appl
40	48	32.0	430	4	US-09-489-039A-12035	Sequence 12035, A
41	48	32.0	493	3	US-09-134-001C-3486	Sequence 3486, Ap
42	48	32.0	724	4	US-08-671-757A-9	Sequence 9, Appl
43	47.5	31.7	308	4	US-09-549-848B-37	Sequence 37, Appl
44	47.5	31.7	308	4	US-09-560-761-10	Sequence 10, Appl
45	47.5	31.7	1065	4	US-09-221-013A-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-097-889-17
; Sequence 17, Application US/09097889
; Patent No. 6218117
; GENERAL INFORMATION:
; APPLICANT: Herinstdt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,889
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman Ph.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 660088.417
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-097-889-17

Query Match 68.3%; Score 102.5; DB 3; Length 68;
Best Local Similarity 69.0%; Pred. No. 7.4e-07;
Matches 20; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

Qy 2 RLNTTWPITPILITLFLITNRLITR 29
Db 3 QLNTTWPITPILITLFLITLQKMENT 31


```
Db      74 LWLLLIPIFUTIFLIAR 92
      :| :: || || || || ||
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42198
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-42198

Query Match      34.3%; Score 51.5; DB 4; Length 456;
Best Local Similarity 30.8%; Pred. No. 31;
Matches 12; Conservative 6; Mismatches 8; Indels 13; Gaps 1;

QY      2 RLNTTVW-----PTIITPILTLFLITNRLI 27
      |||::|||::|||::|||::|||::|||
Db      37 RRNPSMMPLMIHQRLSRHQVPLLIHPLWLTCTCYVTNRRI 75
      |||::|||::|||::|||::|||::|||

RESULT 9
US-09-248-796A-26751
; Sequence 26751, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 26751
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-26751

Query Match      34.0%; Score 51; DB 4; Length 60;
Best Local Similarity 56.2%; Pred. No. 4.6;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY      3 LNTTVWPTIITPILTL 18
      |||::|||::|||::|||::|||
Db      3 LKTLTPTLVVPLVLT 18
      |||::|||::|||::|||::|||

RESULT 10
US-09-543-681A-8270
; Sequence 8270, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8270
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8270

Query Match      34.0%; Score 51; DB 4; Length 302;

Db      74 LWLLLIPIFUTIFLIAR 92
      :| :: || || || || || ||
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 547
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...286
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-547

Query Match      34.7%; Score 52; DB 4; Length 286;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 13; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY      1 LRLNTTVWPTIITPI--LLTLFLITN 24
      |||::|||::|||::|||::|||
Db      14 LCLQTKWSPISPIKAFLYLFVVTN 39
      |||::|||::|||::|||::|||

RESULT 7
US-09-270-767-57472
; Sequence 57472, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57472
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-57472

Query Match      34.3%; Score 51.5; DB 4; Length 263;
Best Local Similarity 30.8%; Pred. No. 18;
Matches 12; Conservative 6; Mismatches 8; Indels 13; Gaps 1;

QY      2 RLNTTVW-----PTIITPILTLFLITNRLI 27
      |||::|||::|||::|||::|||
Db      37 RRNPSMMPLMIHQRLSRHQVPLLIHPLWLTCTCYVTNRRI 75
      |||::|||::|||::|||::|||

RESULT 8
US-09-270-767-42198
; Sequence 42198, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
```

Best Local Similarity 48.1%; Pred. No. 24;
Matches 13; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 LRLNTTVPITPILLFLITRLI 27
||| : : : : :
Db 39 LREGTVRWALIPLLLLLFIILGSLI 65

RESULT 11

US-09-248-796A-20246
; Sequence 20246, Application US/09248796A
; Patent No. 6747137

GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20246
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20246

Query Match 33.3%; Score 50; DB 4; Length 129;
Best Local Similarity 47.6%; Pred. No. 14;
Matches 10; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 LNTTVPITPILLFLIT 23
::: : : : :
Db 6 ISTTKLSVTELLLLFT 26

RESULT 12

US-08-055-917-14
; Sequence 14, Application US/08055917
; Patent No. 5310875

GENERAL INFORMATION:

; APPLICANT: Chang, Tse Wen; Chang, Nancy T.
; TITLE OF INVENTION: Peptides corresponding to membrane-bound Iga
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tanox Biosystems, Inc.
; STREET: 10301 Stella Link Rd.
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77025

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS 3.30
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/055,917
; FILING DATE:
; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/788,120
; FILING DATE: 11/4/1991
; APPLICATION NUMBER: 07/455,080
; FILING DATE: 12/22/1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirabel, Eric P.
; REGISTRATION NUMBER: 31,211
; REFERENCE/DOCKET NUMBER: TNX89-04CCC
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (713) 664-2288
; TELEFAX: (713) 664-8914
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: amino acid
; TOPOLOGY: Linear
US-08-055-917-14

Query Match 32.7%; Score 49; DB 1; Length 65;
Best Local Similarity 64.7%; Pred. No. 9.2;
Matches 11; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 7 VWPITPILLFLIT 23
: : : : :
Db 27 LWPITPILLFLIS 41

RESULT 13

US-08-055-917-16
; Sequence 16, Application US/08055917
; Patent No. 5310875

GENERAL INFORMATION:

; APPLICANT: Chang, Tse Wen; Chang, Nancy T.
; TITLE OF INVENTION: Peptides corresponding to membrane-bound Iga
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tanox Biosystems, Inc.
; STREET: 10301 Stella Link Rd.
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77025

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS 3.30
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/055,917
; FILING DATE:
; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/788,120
; FILING DATE: 11/4/1991
; APPLICATION NUMBER: 07/455,080
; FILING DATE: 12/22/1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirabel, Eric P.
; REGISTRATION NUMBER: 31,211
; REFERENCE/DOCKET NUMBER: TNX89-04CCC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 664-2288
; TELEFAX: (713) 664-8914

; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: amino acid
; TOPOLOGY: Linear
US-08-055-917-16

Query Match 32.7%; Score 49; DB 1; Length 65;
Best Local Similarity 64.7%; Pred. No. 9.2;
Matches 11; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 7 VWPITPILLFLIT 23
: : : : :
Db 27 LWPITPILLFLIS 41

RESULT 14

US-08-055-917-18
; Sequence 18, Application US/08055917

```

; Patent No. 5310875
; GENERAL INFORMATION:
; APPLICANT: Chang, Tse Wen; Chang, Nancy T.
; TITLE OF INVENTION: Peptides corresponding to membrane-bound IgA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tanox Biosystems, Inc.
; STREET: 10301 Stella Link Rd.
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS 3.30
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/055,917
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/788,120
; FILING DATE: 11/4/1991
; APPLICATION NUMBER: 07/455,080
; FILING DATE: 12/22/1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirabel, Eric P.
; REGISTRATION NUMBER: 31,211
; REFERENCE/DOCKET NUMBER: TXN89-04CCC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 664-2288
; TELEFAX: (713) 664-8914
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: amino acid
; TOPOLOGY: Linear
; US-08-055-917-18

Query Match 32.7%; Score 49; DB 1; Length 65;
Best Local Similarity 55.6%; Pred. No. 9.2;
Matches 10; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

Qy 6 TWPTTITPILLTFLIT 23
Db 26 SLWPTTVT--FLTLFLIS 41

RESULT 15
US-08-095-068-14
; Sequence 14, Application US/08095068
; Patent No. 5362643
; GENERAL INFORMATION:
; APPLICANT: Chang, Tse Wen; Chang, Nancy T.
; TITLE OF INVENTION: Producing antibodies which bind to membrane-bound IgA using Ig
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tanox Biosystems, Inc.
; STREET: 10301 Stella Link Rd.
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS 3.30
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,068
; FILING DATE:

```

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/760,765
; FILING DATE: 9/16/1991
; APPLICATION NUMBER: 07/455,080
; FILING DATE: 12/22/1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirabel, Eric P.
; REGISTRATION NUMBER: 31,211
; REFERENCE/DOCKET NUMBER: TXN89-04DEE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 664-2288
; TELEFAX: (713) 664-8914
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: amino acid
; TOPOLOGY: Linear
; US-08-095-068-14

Query Match 32.7%; Score 49; DB 1; Length 65;
Best Local Similarity 64.7%; Pred. No. 9.2;
Matches 11; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

Qy 7 VWPTTITPILLTFLIT 23
Db 27 LWPTTIT--FLTLFLIS 41

Search completed: November 10, 2004, 13:44:01
Job time: 10.3627 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 32.228 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-27

Perfect score: 150

Sequence: 1 LRLLTTVWPTITPILLTLFLITNRLITTR 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225986 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	100.0	30	14	US-10-092-750-27
2	103.5	69.0	65	15	US-10-264-049-4146
3	102.5	68.3	68	9	US-09-098-079-17
4	102.5	68.3	68	16	US-10-408-765A-139
5	98.5	65.7	86	16	US-10-408-765A-1491
6	60.5	40.3	85	17	US-10-425-115-256252
7	55	36.7	75	15	US-10-424-599-187835
8	54	36.0	441	16	US-10-474-776-245
9	53	35.3	81	15	US-10-424-599-159528
10	53	35.3	379	16	US-10-437-963-111334
11	53	35.3	430	14	US-10-369-493-15462
12	53	35.3	430	14	US-10-369-493-15831
13	53	35.3	430	14	US-10-369-493-16212

14	52	34.7	43	17	US-10-425-115-337725	Sequence 337725,
15	52	34.7	286	15	US-10-289-762-547	Sequence 547, App
16	51.5	34.3	131	15	US-10-424-599-169176	Sequence 169176,
17	51	34.0	53	14	US-10-106-698-5361	Sequence 5361, Ap
18	51	34.0	58	14	US-10-106-698-6592	Sequence 6592, Ap
19	51	34.0	337	9	US-09-946-143-2	Sequence 2, Appli
20	51	34.0	337	9	US-09-738-626-3997	Sequence 3997, Ap
21	51	34.0	337	16	US-10-781-014-728	Sequence 728, App
22	51	34.0	648	15	US-10-424-599-276546	Sequence 276546,
23	50.5	33.7	102	15	US-10-424-599-276755	Sequence 276755,
24	50.5	33.7	107	15	US-10-424-599-143296	Sequence 143296,
25	50.5	33.7	557	16	US-10-437-963-193387	Sequence 193387,
26	50.5	33.7	701	9	US-09-900-237-32	Sequence 32, Appl
27	50	33.3	24	14	US-10-195-730-134	Sequence 134, App
28	50	33.3	24	16	US-10-799-747-134	Sequence 134, App
29	50	33.3	257	15	US-10-371-592-1	Sequence 1, Appli
30	49.5	33.0	1377	14	US-10-205-342-25	Sequence 25, Appl
31	49	32.7	117	15	US-10-424-599-272186	Sequence 272186,
32	49	32.7	226	9	US-09-098-079-18	Sequence 18, Appl
33	49	32.7	299	15	US-10-282-923A-53880	Sequence 53880, A
34	49	32.7	306	16	US-10-437-963-126321	Sequence 126321,
35	49	32.7	407	16	US-10-437-963-129466	Sequence 129466,
36	48.5	32.3	94	17	US-10-435-115-194741	Sequence 194741,
37	48.5	32.3	247	15	US-10-424-599-281162	Sequence 281162,
38	48.5	32.3	446	15	US-10-617-038-8	Sequence 8, Appli
39	48	32.0	44	15	US-10-424-599-280251	Sequence 280251,
40	48	32.0	106	15	US-10-264-237-1613	Sequence 1613, Ap
41	48	32.0	115	16	US-10-437-963-189602	Sequence 189602,
42	48	32.0	173	16	US-10-437-963-136989	Sequence 136989,
43	48	32.0	198	14	US-10-029-386-31990	Sequence 31990, A
44	48	32.0	226	16	US-10-408-765A-192	Sequence 192, App
45	48	32.0	259	9	US-09-996-634-133	Sequence 133, App

ALIGNMENTS

RESULT 1
US-10-092-750-27
; Sequence 27, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-27

Query Match 100.0%; Score 150; DB 14; Length 30;
Best Local Similarity 100.0%; Pred. No.: 2.4e-13;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LRLLTTVWPTITPILLTLFLITNRLITTR 30
DB 1 LRLLTTVWPTITPILLTLFLITNRLITTR 30
RESULT 2
US-10-264-049-4146
; Sequence 4146, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:

Dbb 27 TKAWPTFISYLLSVLFMIVN 46

; APPLICANT: Barbazuk, Brad

APPLICANT: Boukharov, Andrey A.

```
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 111334
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_15321C.1.pap
US-10-437-963-111334

Query Match      35.3%; Score 53; DB 16; Length 379;
Best Local Similarity 40.0%; Pred. No. 50;
Matches 10; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY      3 LNTTVPPTIITPILTLFLITNRLI 27
DB      254 LPTTHWPTIITMAVLLSILSSFL 278

RESULT 11
US-10-369-493-15462
; Sequence 15462, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15462
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15462

Query Match      35.3%; Score 53; DB 14; Length 430;
Best Local Similarity 40.0%; Pred. No. 57;
Matches 10; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY      3 LNTTVPPTIITPILTLFLITNRLI 27
DB      103 LLATVWPALPVPMLASVFMATFTVI 127

RESULT 12
US-10-369-493-15831
; Sequence 15831, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
```

```
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15831
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15831

Query Match      35.3%; Score 53; DB 14; Length 430;
Best Local Similarity 40.0%; Pred. No. 57;
Matches 10; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY      3 LNTTVPPTIITPILTLFLITNRLI 27
DB      103 LLATVWPALPVPMLASVFMATFTVI 127

RESULT 13
US-10-369-493-16212
; Sequence 16212, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16212
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16212

Query Match      35.3%; Score 53; DB 14; Length 430;
Best Local Similarity 40.0%; Pred. No. 57;
Matches 10; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY      3 LNTTVPPTIITPILTLFLITNRLI 27
DB      103 LLATVWPALPVPMLASVFMATFTVI 127

RESULT 14
US-10-425-115-337725
; Sequence 337725, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 337725
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_71175C.1.pap
US-10-425-115-337725

Query Match      34.7%; Score 52; DB 17; Length 43;
```

Search completed: November 11, 2004, 01:28:09
Job time : 33.278 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 6.68394 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-27

Perfect score: 150

Sequence: 1 LRLNTTVMPTIITPILTLFLITNRLITR 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102.5	68.3	68	1	PHU8
2	97	64.7	68	2	E9183
3	97	64.7	68	2	T14024
4	96	64.0	68	2	T14197
5	95	63.3	68	2	T14147
6	90	60.0	68	2	I37051
7	83	55.3	68	2	T14143
8	79	52.7	68	2	T11837
9	57	38.0	68	2	T11510
10	56	37.3	169	2	Al1421
11	54	36.0	66	1	PWBO8
12	54	36.0	395	2	D37155
13	54	36.0	440	2	H95054
14	54	36.0	441	2	E97924
15	53	35.3	534	2	A9316
16	53	35.3	625	2	H98148
17	53	35.3	651	2	A93139
18	52.5	35.0	336	2	T25180
19	52	34.7	469	2	F97738
20	51.5	34.3	166	2	I37052
21	51.5	34.3	226	1	PWBO6
22	51.5	34.3	226	2	T11838
23	51.5	34.3	650	2	A94976
24	51	34.0	881	2	A92777
25	51	34.0	881	2	B37557
26	50	33.3	67	2	T11144
27	49.5	33.0	1254	2	T47141
28	49	32.7	66	2	T11054
29	49	32.7	67	2	T10976

30	49	32.7	181	2	H72557	hypothetical prote
31	49	32.7	226	1	PHU6	H+-transporting tw
32	49	32.7	305	2	AC2749	ABC transporter, m
33	49	32.7	305	2	B97530	trehalose/maltose
34	49	32.7	325	2	D89974	conserved hypothet
35	48.5	32.3	446	2	B70543	hypothetical prote
36	48	32.0	63	2	E58850	H+-transporting tw
37	48	32.0	67	2	S41839	H+-transporting tw
38	48	32.0	67	2	T11484	H+-transporting tw
39	48	32.0	261	2	E70957	probable lbrf prot
40	48	32.0	352	2	D85826	probable transport
41	48	32.0	352	2	D64966	membrane protein y
42	48	32.0	352	2	G90980	probable transport
43	48	32.0	477	2	AC0122	Sodium/galactoside
44	48	32.0	724	2	C81361	flagellar biosynth
45	48	32.0	724	2	A49217	flagellar biogenes

ALIGNMENTS

RESULT 1

PMU8

H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - human mitochondrion

N;Alternate names: hydrogen ion-transporting ATP synthase protein 8

C;Species: mitochondrion Homo sapiens (man)

C;Date: 22-May-1981 #sequence_revision 23-Oct-1981 #text_change 09-Jul-2004

C;Accession: A01062; I80237

R;Anderson, S.; Bankier, A.T.; Barrell, B.G.; de Bruijn, M.H.L.; Coulson, A.R.; Drouin, J.

Nature 290, 457-465, 1981

A;Title: Sequence and organization of the human mitochondrial genome.

A;Reference number: A00151; MUID:81173052; PMID:7219534

A;Accession: A01062

A;Molecule type: DNA

A;Residues: 1-68 <AND>

A;Cross-references: UNIPROT:P03928; GB:J01415; GB:M12548; GB:M58503; GB:M63932; GB:M63933

R;Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.

Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995

A;Title: Recent African origin of modern humans revealed by complete sequences of hominoid

A;Reference number: I59384; MUID:95132634; PMID:7530363

A;Accession: I80237

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-31, 'S', '33-68 <RES>

A;Cross-references: GB:D38112; NID:9644480; PIDN:BAA07294.1; PID:9704445

C;Genetics:

A;Gene: GDB:MTATP8

A;Cross-references: GDB:118898; OMIM:516070

A;Map position: MTH8366-8572

A;Genome: mitochondrion

A;Genetic code: SGCI

C;Superfamily: H+-transporting ATP synthase protein 8

C;Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxid

Query Match 68.3%; Score 102.5; DB 1; Length 68;
Best Local Similarity 69.0%; Pred. No. 3.9e-07;
Matches 20; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

Qy 2 RLNTTVMPTIITPILTLFLITN-RLITT 29

Db 3 QLNTTVMPTMITPMLLTFLITQLKMLNT 31

RESULT 2

E9153

H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - western lowland gorilla mitoc

N;Alternate names: hydrogen ion-transporting ATP synthase protein 8

C;Species: mitochondrion Gorilla gorilla (western lowland gorilla)

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Jun-2002

C;Accession: E59153

R;Xu, X.; Arnason, U.

Mol. Biol. Evol. 13, 691-698, 1996

A:Title: A complete sequence of the mitochondrial genome of the Western lowland gorilla.
A:Reference number: Z17269; MUID:96212991; PMID:9676744
A:Accession: E59153
A>Status: preliminary; nucleic acid sequence not shown; translation not shown; translate
A:Molecule type: DNA
A:Residues: 1-68 <XUX>
A:CROSS-references: GB:Y93347; NID:G1304307; GSPDB:GN00106
A>Note: submitted to GenBank, November 1995
A:Note: this translation is not annotated in GenBank entry GGMITG, release 114.0
C:Genetics:
A:Gene: ATPase8
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: H+-transporting ATP synthase protein 8
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxil

Query Match 64.7%; Score 97; DB 2; Length 68;
Best Local Similarity 81.8%; Pred. No. 2.2e-06;
Matches 18; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLNTTWPPTIIPILLTLFLIT 23
:|||||:|:|||||
Db 3 QLNTTWPPTIIPMLLTFLIT 24

RESULT 3
T14024
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - gorilla mitochondrion
C:Species: mitochondrion Gorilla gorilla (gorilla)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T14024
R:Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.
Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995
A:Title: Recent African origin of modern humans revealed by complete sequences of hominid
A:Reference number: I59384; MUID:95132634; PMID:7530363
A:Accession: T14024
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-68 <HOR>
A:CROSS-references: UNIPROT:Q34571; EMBL:D38114; PIDN:BAA07304.1
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: H+-transporting ATP synthase protein 8
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxil

Query Match 64.7%; Score 97; DB 2; Length 68;
Best Local Similarity 81.8%; Pred. No. 2.2e-06;
Matches 18; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLNTTWPPTIIPILLTLFLIT 23
:|||||:|:|||||
Db 3 QLNTTWPPTIIPMLLTFLIT 24

RESULT 4
T14197
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - chimpanzee mitochondrion
C:Species: mitochondrion Pan troglodytes (chimpanzee)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T14197
R:Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.
Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995
A:Title: Recent African origin of modern humans revealed by complete sequences of hominid
A:Reference number: I59384; MUID:95132634; PMID:7530363
A:Accession: T14197
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-68 <HOR>
A:CROSS-references: UNIPROT:Q35647; EMBL:D38113; PIDN:BAA07300.1
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1

C:Superfamily: H+-transporting ATP synthase protein 8
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxil

Query Match 64.0%; Score 96; DB 2; Length 68;
Best Local Similarity 77.3%; Pred. No. 3e-06;
Matches 17; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLNTTWPPTIIPILLTLFLIT 23
:|||||:|:|||||
Db 3 QLNTTWPPTIIPMLLTFLIT 24

RESULT 5
T14147
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - pygmy chimpanzee mitochondrion
C:Species: mitochondrion Pan paniscus (pygmy chimpanzee, bonobo)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T14147
R:Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.
Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995
A:Title: Recent African origin of modern humans revealed by complete sequences of hominid
A:Reference number: I59384; MUID:95132634; PMID:7530363
A:Accession: T14147
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-68 <HOR>
A:CROSS-references: UNIPROT:Q35587; EMBL:D38116; PIDN:BAA07313.1
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: H+-transporting ATP synthase protein 8
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxil

Query Match 63.3%; Score 95; DB 2; Length 68;
Best Local Similarity 81.8%; Pred. No. 4e-06;
Matches 18; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLNTTWPPTIIPILLTLFLIT 23
:|||||:|:|||||
Db 3 QLNTTWPPTIIPMLLTFLIT 24

RESULT 6
I37051
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - siamang mitochondrion
C:Species: mitochondrion Hylobates syndactylus (siamang)
C:Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 09-Jul-2004
C:Accession: I37051
R:Horai, S.; Satta, Y.; Hayasaka, K.; Kondo, R.; Inoue, T.; Ishida, T.; Hayashi, S.; Taka
J. Mol. Evol. 35, 32-43, 1992
A:Title: Man's place in Hominoida revealed by mitochondrial DNA genealogy.
A:Reference number: I37047; MUID:92389366; PMID:1518083
A:Accession: I37051
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-68 <HOR>
A:CROSS-references: UNIPROT:Q34801; GB:D38484; NID:9558513; PIDN:BAA07498.1; PID:9558518
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: H+-transporting ATP synthase protein 8
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxil

Query Match 60.0%; Score 90; DB 2; Length 68;
Best Local Similarity 81.0%; Pred. No. 1.9e-05;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 RLNTTWPPTIIPILLTLFLI 22
:|||||:|:|||||
Db 3 QLNTTWPPTIIPMLLTFLFL 23

RESULT 7

Tl14143
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - orangutan mitochondrion
C:Species: mitochondrion Pongo pygmaeus (orangutan)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: Tl14143
R:Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.
Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995
A:Title: Recent African origin of modern humans revealed by complete sequences of hominid cytochrome b
A:Reference number: 159384; MUID:95132634; PMID:7530363
A:Accession: Tl14143
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-68 <HOR>
A:Cross-references: UNIPROT:Q35584; EMBL:D38115; PIDN:BA07309.1
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: H+-transporting ATP synthase protein 8
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxidative phosphorylation

Query Match 55.3%; Score 83; DB 2; Length 68;
Best Local Similarity 72.7%; Pred. No. 0.00017;
Matches 16; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 RLNTTWPPTIITPILLTLFLIT 23
:||||| :||||| :|||||
Db 3 QLNTTTLTWTPTLLALFLIT 24
:||||| :||||| :|||||

RESULT 8
Tl1837
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - common gibbon mitochondrion
C:Species: mitochondrion Hylobates lar (common gibbon, white-handed gibbon)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: Tl1837
R:Arnason, U.; Gullberg, A.; Xu, X.
Hereditas 124, 185-189, 1996
A:Title: A complete mitochondrial DNA molecule of the white-handed gibbon, Hylobates lar
A:Reference number: 217353
A:Accession: Tl1837
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-68 <ARN>
A:Cross-references: UNIPROT:Q95705; EMBL:X99256; PIDN:CAA67632.1
A:Experimental source: isolate Ester
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: H+-transporting ATP synthase protein 8
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxidative phosphorylation

Query Match 52.7%; Score 79; DB 2; Length 68;
Best Local Similarity 71.4%; Pred. No. 0.00059;
Matches 15; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 RLNTTWPPTIITPILLTLFLI 22
:||||| :||||| :|||||
Db 3 QLNTTWPPTIITPILLTLFLI 23
:||||| :||||| :|||||

RESULT 9
Tl1510
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - baboon mitochondrion
C:Species: mitochondrion Papio hamadryas (baboon)
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: Tl1510
R:Arnason, U.; Gullberg, A.; Janke, A.
J. Mol. Evol. 47, 718-727, 1998
A:Title: Molecular timing of primate divergences as estimated by two non-primate calibrations
A:Reference number: 217277; MUID:99065765; PMID:9847414
A:Accession: Tl1510
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-68 <ARN>
A:Cross-references: UNIPROT:Q92XY0; EMBL:Y18001; NID:94049475; PIDN:CAA76998.1; PID:94045;
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: H+-transporting ATP synthase protein 8
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxidative phosphorylation

Query Match 38.0%; Score 57; DB 2; Length 68;
Best Local Similarity 54.5%; Pred. No. 0.056;
Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 RLNTTWPPTIITPILLTLFLIT 23
:||||| :||||| :|||||
Db 3 QLDTSTWFTIIMWMLPTLYLIT 24
:||||| :||||| :|||||

RESULT 10
Al1421
hypothetical protein lmo2778 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: Al1421
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Ftschi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Matok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: Al1421
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-169 <GLA>
A:Cross-references: UNIPROT:Q9Y305; GB:NC_003210; PIDN:CAD00991.1; PID:916412278; GSPDB:1
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2778

Query Match 37.3%; Score 56; DB 2; Length 169;
Best Local Similarity 48.0%; Pred. No. 1.9;
Matches 12; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 3 LNTTWPPTIITPILLTLFLIINRLI 27
:||||| :||||| :|||||
Db 91 INRNITATIIIGIITLFLIAGTII 115
:||||| :||||| :|||||

RESULT 11
P0808
H+-transporting two-sector ATPase (EC 3.6.3.14) protein 8 - bovine mitochondrion
N:Alternate names: A6L protein; hydrogen ion-transporting ATP synthase protein 8
C:Species: mitochondrion Bos primigenius taurus (cattle)
C:Date: 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change 09-Jul-2004
C:Accession: A01063; A25474; D39566
R:Anderson, S.; de Bruijn, M.H.L.; Coulson, A.R.; Eperon, I.C.; Sanger, F.; Young, I.G.
J. Mol. Biol. 156, 683-717, 1982
A:Title: Complete sequence of bovine mitochondrial DNA. Conserved features of the mammal
A:Reference number: A00152; MUID:83010260; PMID:7120390
A:Accession: A01063
A:Molecule type: DNA
A:Residues: 1-66 <AND>
A:Cross-references: UNIPROT:P03929; GB:J01394; NID:9336430; PIDN:AAB59272.1; PID:9336435;
R:Pearlley, I.M.; Walker, J.E.
EMBO J. 5, 2003-2008, 1986
A:Title: Two overlapping genes in bovine mitochondrial DNA encode membrane components of
A:Reference number: A25474; MUID:87004570; PMID:2875870
A:Accession: A25474
A:Molecule type: protein
A:Residues: 1-37; 45-50; 52-53 <FEA>
R:Walker, J.E.; Lutter, R.; Dupuis, A.; Runswick, M.J.
Biochemistry 30, 5369-5378, 1991

Job time : 7.68394 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 36.6321 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-27

Perfect score: 150

Sequence: 1 LRNTTWPPIITPILLTLELTNRLITR 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	102.5	68.3	68	1	ATP8_HUMAN
2	102.5	68.3	68	2	Q6R326
3	102.5	68.3	68	2	Q6R326
4	102.5	68.3	68	2	Q6R326
5	102.5	68.3	68	2	Q6R326
6	102.5	68.3	68	2	Q6R326
7	102.5	68.3	68	2	Q6R326
8	102.5	68.3	68	2	Q6R326
9	102.5	68.3	68	2	Q6R326
10	102.5	68.3	68	2	Q6R326
11	102.5	68.3	68	2	Q6R326
12	102.5	68.3	68	2	Q6R326
13	102.5	68.3	68	2	Q6R326
14	102.5	68.3	68	2	Q6R326
15	102.5	68.3	68	2	Q6R326
16	102.5	68.3	68	2	Q6R326
17	102.5	68.3	68	2	Q6R326
18	102.5	68.3	68	2	Q6R326
19	102.5	68.3	68	2	Q6R326
20	102.5	68.3	68	2	Q6R326
21	102.5	68.3	68	2	Q6R326
22	102.5	68.3	68	2	Q6R326
23	102.5	68.3	68	2	Q6R326
24	102.5	68.3	68	2	Q6R326
25	102.5	68.3	68	2	Q6R326
26	102.5	68.3	68	2	Q6R326
27	102.5	68.3	68	2	Q6R326
28	102.5	68.3	68	2	Q6R326
29	102.5	68.3	68	2	Q6R326
30	102.5	68.3	68	2	Q6R326
31	102.5	68.3	68	2	Q6R326

RESULT 1

ATP8_HUMAN STANDARD; PRT; 68 AA.

AC P03928; Q34771;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).

GN Name=MTATP8; Synonyms=ATP8;

OS Homo sapiens (Human).

OC Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]_TaxID=9606;

RP SEQUENCE FROM N.A.

RX MEDLINE=81173052; PubMed=7219534;

RA Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L.,

RA Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A.,

RA Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G.;

RT "Sequence and organization of the human mitochondrial genome.";

RL Nature 290:457-465(1981).

RN [2]_TaxID=9606;

RP SEQUENCE FROM N.A.

RX TISSUE=Placenta;

RC MEDLINE=95132634; PubMed=7530363;

RA Horai S., Hayasaka K., Kondo R., Tsugane K., Takahata N.;

RT "Recent African origin of modern humans revealed by complete sequences

of hominoid mitochondrial DNAs.";

RL Proc. Natl. Acad. Sci. U.S.A. 92:532-536(1995).

RN [3]_TaxID=9606;

RP SEQUENCE FROM N.A.

RX PubMed=12949126; DOI=10.1093/molbev/msg230;

RA McIlanen J.S., Finnila S., Majamaa K.;

RT "Lineage-Specific Selection in Human mtDNA: Lack of Polymorphisms in a

Segment of MTND5 Gene in Haplogroup J.";

RL Mol. Biol. Evol. 20:2132-2142(2003).

RN [4]_TaxID=9606;

RP SEQUENCE FROM N.A.

RX MEDLINE=21012010; PubMed=11130070; DOI=10.1038/35047064;

RA Ingman M., Kaessmann H., Paabo S., Gyllenstein U.;

RT "Mitochondrial genome variation and the origin of modern humans.";

RL Nature 408:708-713(2000).

RN [5]_TaxID=9606;

RP SEQUENCE FROM N.A.

RX MEDLINE=22723755; PubMed=12840039; DOI=10.1101/gr.686603;

RA Ingman M., Gyllenstein U.;

RT "Mitochondrial genome variation and evolutionary history of Australian

and new guinean aborigines.";

RL Genome Res. 13:1600-1606(2003).

RN [6]_TaxID=9606;

RP SEQUENCE FROM N.A.

RX PubMed=14760490; DOI=10.1007/s00414-004-0427-6;

RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.;

RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RL increase the power of forensic testing in Caucasians";
RN Int. J. Legal Med. 118:137-146(2004).
[7]
RP VARIANTS PRO-17 AND SER-21.
RX MEDLINE=9209084; PubMed=1757091;
RA Marzuki S., Noer A.S., Lertit P., Thyagarajan D., Kapsa R.,
RA Utthanaphol P., Byrne E.;
RT "Normal variants of human mitochondrial DNA and translation products:
RL the building of a reference data base";
RN Hum. Genet. 88:139-145(1991).
[8]
RP VARIANT THR-28.
RX MEDLINE=98127994; PubMed=9461455;
RA Rieder M.J., Taylor S.L., Tobe V.O., Nickerson D.A.;
RT "Automating the identification of DNA variations using quality-based
RT fluorescence re-sequencing: analysis of the human mitochondrial
RT genome";
RL Nucleic Acids Res. 26:967-973(1998).
CC -!- FUNCTION: This is one of the chains of the nonenzymatic component
CC (CF0) subunit of the mitochondrial ATPase complex.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBCELLULAR LOCATION: Membrane-bound.
CC -!- SIMILARITY: Belongs to the ATPase protein 8 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J01415; AAB58947.1; -
DR EMBL; V00662; CRA24030.1; -
DR EMBL; D38112; BAA07294.1; -
DR EMBL; AY339402; AAP89040.1; -
DR EMBL; AY339403; AAP89053.1; -
DR EMBL; AY339404; AAP89066.1; -
DR EMBL; AY339405; AAP89079.1; -
DR EMBL; AY339406; AAP89092.1; -
DR EMBL; AY339407; AAP89105.1; -
DR EMBL; AY339408; AAP89118.1; -
DR EMBL; AY339409; AAP89131.1; -
DR EMBL; AY339410; AAP89144.1; -
DR EMBL; AY339411; AAP89157.1; -
DR EMBL; AY339412; AAP89170.1; -
DR EMBL; AY339413; AAP89183.1; -
DR EMBL; AY339414; AAP89196.1; -
DR EMBL; AY339415; AAP89209.1; -
DR EMBL; AY339416; AAP89222.1; -
DR EMBL; AY339417; AAP89235.1; -
DR EMBL; AY339418; AAP89248.1; -
DR EMBL; AY339419; AAP89261.1; -
DR EMBL; AY339420; AAP89274.1; -
DR EMBL; AY339421; AAP89287.1; -
DR EMBL; AY339422; AAP89300.1; -
DR EMBL; AY339423; AAP89313.1; -
DR EMBL; AY339424; AAP89326.1; -
DR EMBL; AY339425; AAP89339.1; -
DR EMBL; AY339426; AAP89352.1; -
DR EMBL; AY339427; AAP89365.1; -
DR EMBL; AY339428; AAP89378.1; -
DR EMBL; AY339429; AAP89391.1; -
DR EMBL; AY339430; AAP89404.1; -
DR EMBL; AY339431; AAP89417.1; -
DR EMBL; AY339432; AAP89430.1; -
DR EMBL; AY339433; AAP89443.1; -
DR EMBL; AY339434; AAP89456.1; -
DR EMBL; AY339435; AAP89469.1; -
DR EMBL; AY339436; AAP89482.1; -
DR EMBL; AY339437; AAP89495.1; -
DR EMBL; AY339438; AAP89508.1; -
DR EMBL; AY339439; AAP89521.1; -
DR EMBL; AY339440; AAP89534.1; -
DR EMBL; AY339441; AAP89547.1; -
DR EMBL; AY339442; AAP89560.1; -
DR EMBL; AY339443; AAP89573.1; -
DR EMBL; AY339444; AAP89586.1; -
DR EMBL; AY339445; AAP89599.1; -
DR EMBL; AY339446; AAP89612.1; -
DR EMBL; AY339447; AAP89625.1; -
DR EMBL; AY339448; AAP89638.1; -
DR EMBL; AY339449; AAP89651.1; -
DR EMBL; AY339450; AAP89664.1; -
DR EMBL; AY339451; AAP89677.1; -
DR EMBL; AY339452; AAP89690.1; -
DR EMBL; AY339453; AAP89703.1; -
DR EMBL; AY339454; AAP89716.1; -
DR EMBL; AY339455; AAP89729.1; -
DR EMBL; AY339456; AAP89742.1; -
DR EMBL; AY339457; AAP89755.1; -
DR EMBL; AY339458; AAP89768.1; -
DR EMBL; AY339459; AAP89781.1; -
DR EMBL; AY339460; AAP89794.1; -
DR EMBL; AY339461; AAP89807.1; -
DR EMBL; AY339462; AAP89820.1; -
DR EMBL; AY339463; AAP89833.1; -
DR EMBL; AY339464; AAP89846.1; -
DR EMBL; AY339465; AAP89859.1; -
DR EMBL; AY339466; AAP89872.1; -
DR EMBL; AY339467; AAP89885.1; -
DR EMBL; AY339468; AAP89898.1; -
DR EMBL; AY339469; AAP89911.1; -
DR EMBL; AY339470; AAP89924.1; -
DR EMBL; AY339471; AAP89937.1; -
DR EMBL; AY339472; AAP89950.1; -
DR EMBL; AY339473; AAP89963.1; -
DR EMBL; AY339474; AAP89976.1; -
DR EMBL; AY339475; AAP89989.1; -
DR EMBL; AY339476; AAP90002.1; -
DR EMBL; AY339477; AAP90015.1; -
DR EMBL; AY339478; AAP90028.1; -
DR EMBL; AY339479; AAP90041.1; -
DR EMBL; AY339480; AAP90054.1; -
DR EMBL; AY339481; AAP90067.1; -
DR EMBL; AY339482; AAP90080.1; -
DR EMBL; AY339483; AAP90093.1; -
DR EMBL; AY339484; AAP90106.1; -
DR EMBL; AY339485; AAP90119.1; -
DR EMBL; AY339486; AAP90132.1; -
DR EMBL; AY339487; AAP90145.1; -
DR EMBL; AY339488; AAP90158.1; -
DR EMBL; AY339489; AAP90171.1; -
DR EMBL; AY339490; AAP90184.1; -
DR EMBL; AY339491; AAP90197.1; -
DR EMBL; AY339492; AAP90210.1; -
DR EMBL; AY339493; AAP90223.1; -
DR EMBL; AY339494; AAP90236.1; -
DR EMBL; AY339495; AAP90249.1; -
DR EMBL; AY339496; AAP90262.1; -
DR EMBL; AY339497; AAP90275.1; -
DR EMBL; AY339498; AAP90288.1; -
DR EMBL; AY339499; AAP90301.1; -
DR EMBL; AY339500; AAP90314.1; -
DR EMBL; AY339501; AAP90327.1; -
DR EMBL; AY339502; AAP90340.1; -
DR EMBL; AY339503; AAP90353.1; -
DR EMBL; AY339504; AAP90366.1; -
DR EMBL; AY339505; AAP90379.1; -
DR EMBL; AY339506; AAP90392.1; -
DR EMBL; AY339507; AAP90405.1; -
DR EMBL; AY339508; AAP90418.1; -
DR EMBL; AY339509; AAP90431.1; -